

N° attribué par la bibliothèque

□□□□□□□□□□

THESE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE PARIS OUEST NANTERRE LA DEFENSE

Discipline : PSYCHOLOGIE

présentée et soutenue publiquement

par

Mariya Jilinskaya

Octobre 2012

Approche psychométrique et différentielle de la mesure du leadership par la
méthode à 360 degrés.

Artefact et réalité dans l'hétéro-évaluation.

Directeur de thèse : **Pierre VRIGNAUD**

Professeur de Psychologie du Travail et de l'Orientation, Université Paris-Ouest Nanterre
La Défense, EA 4431.

Membres du jury :

Manfred KET DE VRIES The Raoul de Vitry d'Avaucourt Chaired Distinguished Clinical,
Professor of Leadership Development and Organisational
Change, INSEAD.

Claude HOUSSEMAND Professeur de Psychologie Différentielle, Université du
Luxembourg/EMACS (Educational Measurement and Applied
Cognitive Science).

Even LOARER Professeur titulaire de la Chaire de Psychologie de
l'Orientation, CNAM/INETOP, CRTD, EA4132.

Jean-Michel PETOT Professeur de Psychologie, Université Paris-Ouest Nanterre La
Défense, IPSE EA 4432.

Remerciements

Je remercie tout d'abord les membres du jury d'avoir accepté d'évaluer ce travail.

Je remercie tout particulièrement mon directeur de thèse, Pierre VRIGNAUD, pour avoir cru (souvent plus que moi) dans mes capacités à terminer ce travail. Ses nombreuses connaissances et la facilité avec laquelle Monsieur VRIGNAUD les partage m'ont appris énormément durant ces années de doctorat. Enfin, son expertise en psychologie et en statistique, sa bienveillance et son soutien de tous les instants sont ont été une aide et une motivation précieuse pour moi.

Je remercie également l'équipe de l'INSEAD et particulièrement le Professeur KETS DE VRIES et Elisabeth FLORENT pour leur temps et leur disponibilité, mais surtout pour m'avoir fait partager la passion de leur travail et m'avoir si gentiment accueilli dans leur équipe. Je les remercie aussi, de m'avoir permis de travailler sur les données recueillis dans le cadre des formations dispensés au Leadership Center. Cette collaboration reste pour moi une expérience très enrichissante et motivante.

À Carolyn GRANIER-DEFERRE qui a cru en moi depuis ma maîtrise, et a continué à m'encourager depuis, je dis un grand merci. Je lui suis également reconnaissante de m'avoir présenté le Professeur VRIGNAUD et de m'avoir encouragé à faire ma thèse avec lui.

Merci à ma famille, qui m'a soutenue durant tout ce travail de thèse, même lorsque la soutenance semblait s'éloigner plus que se rapprocher. Merci à ma mère, pour ses encouragements et sa patience, et à mon père pour nos discussions constructives et ses *insights* mathématiques et statistiques qui m'ont souvent aidé à regarder les choses différemment.

A tous mes amis, qui n'ont pas directement collaboré à ce travail, mais m'ont permis de penser à autre chose, d'avoir une vie sociale et de me rappeler qu'il n'y a pas que la recherche dans la vie - un grand merci à Perrine, Nathalie, Ana, Madeleine et Céline.

Enfin, merci à Vijai, qui m'a offert du temps pour travailler sereinement sur ma thèse et m'a donné, de par son soutien inconditionnel, les moyens de mener ce travail jusqu'au bout, et tellement plus encore. Cette thèse est aussi la tienne...

Résumé

Cette thèse porte sur l'évaluation du leadership par une approche multi-évaluateurs (dite à 360 degrés). Tout d'abord les différents modèles du leadership, allant d'une conception unitaire, à une conception interactionniste puis à une définition en termes d'effet, sont détaillés. Puis en étudiant la question de la mesure, on met en évidence qu'avec la popularité croissante de leurs modèles, certains outils d'évaluation sont devenus des questionnaires psychométriques à part entière. Pourtant, du fait des limites de l'auto-évaluation, une nouvelle approche du leadership à vu le jour: l'évaluation à 360°. Elle évalue les qualités d'un manager en interrogeant les personnes travaillant avec lui (subordonnés, collègues, supérieurs...) et en comparant leurs évaluations avec la propre évaluation du manager. Un des points central de notre recherche a été d'étudier les apports et les limites de cette méthode. Tout d'abord on a vérifié dans quelle mesure les caractéristiques souvent utilisées pour expliquer la variabilité entre les catégories d'observateurs permettaient réellement de comprendre les écarts observés. Ces analyses ont montré que malgré des résultats significatifs, ces variables n'expliquent que très partiellement la variance existante. De par ces conclusions l'accent a été mis, non plus sur les différences inter-groupes, mais sur l'accord et le désaccord au sein des groupes d'observateurs. Enfin, la dernière partie revient aux bases méthodologiques et théoriques de la mesure en cherchant à proposer un modèle psychométriques qui conviendrait aux résultats de questionnaires à 360°, permettant de donner un cadre conceptuel au recours à des évaluateurs multiples.

Mots clefs : Leadership, 360 degrés, analyses multi-niveaux, accord inter-observateur, psychométrie, variabilité inter et intra-individuelle

Abstract

This thesis is centered on leadership assessment through multi-rater evaluation, commonly known as 360 degrees assessment. First, leadership models were presented, and then, we discussed the measurement aspects of leadership, wherein we observed that some tools became full fledged psychometric assessments owing to the growing popularity of their underlying theory. Nevertheless, the concerns over the inherent limitations of self-report measures continued to be a major challenge in leadership assessment. This led to a new assessment approach called 360 degrees in which the characteristics of leaders are assessed by people working with them (subordinates, colleagues, superiors...) and compared with the leaders' self-appraisals. The focal point of this thesis was to study the advantages as well as the limitations of this approach. The study started with examining how well the variables which are supposed to explain the inter-rater variability were actually helpful in understanding the observed variance among observers. Those analysis yielded significant results despite the fact that those variables could explain only a very limited amount of variance. Following these observations, the study switched its focus from inter-group differences to intra-group / inter-rater agreement and disagreement. Finally, the last part of this thesis gets back to methodological and theoretical basics of measurement theory and proposes a psychometric model that would suit the 360 degrees assessments followed by a conceptual framework for the studies using multi-rater techniques.

Key words: Leadership, 360 degrees, multi-level analysis, inter-rater agreement, psychometrics, inter and intra-individual variability.

Table des matières

Remerciements.....	2
Résumé.....	3
Abstract.....	4
Introduction	12
Chapitre 1. Approche historique du concept de leadership.....	15
1. <i>Les théories unitaires.....</i>	<i>18</i>
A. Les traits de personnalité	18
1. Great Man Theory.....	18
2. Le leadership comme un construit secondaire des tests de personnalité	19
B. Les théories Environnementales	20
C. Les approches comportementales : les comportements et les styles du leader....	20
1. Les études de l'Université d'Iowa.....	21
A. Lewin, Lippit et White (1939).....	21
B. Le continuum de Tannenbaum et Schmidt (1958).....	22
2. Théorie X et Théorie Y.....	23
3. Les études de l'Université du Michigan.....	24
4. Les études de l'Université de l'Ohio	26
5. Les travaux de Blake et Mouton.....	27
6. Les limites et critiques de l'approche du leadership sous forme de style	28
2. <i>L'intégration des variables situationnelles et de contingence.....</i>	<i>30</i>
A. La théorie de Contingence de Fiedler (1967)	30
B. Le modèle en Trois Dimensions de Reddin (1986)	32
C. Le leadership situationnel, Hersey et Blanchard (1969)	34
D. La théorie « trajectoire-but » (Path-goal Theory).....	35
E. Le modèle Normatif de Vroom et Yetton (1973), Vroom et Jago (1988).....	37
3. <i>Les nouvelles approches du Leadership.....</i>	<i>40</i>
A. Le leader charismatique.....	40
1. La théorie de House (1977) du leader charismatique.....	40
2. La théorie de l'attribution du Charisme de Conger et Kanungo (1987). 43	

3. Théorie du Self-Concept de Shamir (1991; 1993).....	44
B. Le leadership transformationnel.....	46
1. La théorie des échanges Leader – Suiveur	46
2. Le leadership Moral de Burns (1978)	46
3. Le leadership transformationnel de Bass (1985; 1987).....	48
4. James Kouzes et Barry Posner (1987) : les meilleurs comportements de leadership.....	49
5. Comparaison entre les modèles	50
6. La théorie Intégratrice du leadership de Chemers (1997).....	51
5. L'approche clinique du leadership.....	52

Chapitre 2 : Le Leadership, une dimension pas si indépendante.....56

1. Les variables d'effets et l'efficacité du leadership	56
A. Perception de l'efficacité des leaders.....	57
B. Efficacité et perception de l'efficacité du groupe.....	58
C. Satisfaction des employés avec le leader / le travail.....	59
D. Efforts supplémentaires des employés / implication dans le travail.....	60
E. Intention de quitter le poste.....	61
F. Engagement envers l'organisation	62
2. Lien avec d'autres dimensions psychologiques.....	63
A. Les traits de personnalité.....	63
1. Les cinq dimensions fondamentales de la personnalité.....	64
2. Liens Big Five - Leadership.....	66
3. Le modèle HEXACO de la personnalité.....	68
4. Les types de personnalité.....	69
B L'intelligence émotionnelle.....	71

Chapitre 3. La mesure du leadership.....75

1. La mesure des qualités managériales et de leadership	76
A. Organization Description Questionnaire, House et Rizzo (1972).....	76
B. Managerial Practices Survey de Yukl, (1990).....	77
2. Les mesures du type et du style de leader	79
A. Le Leadership Behaviour Description Questionnaire	79
B. Le Leader Opinion Questionnaire.....	80

C. Managerial Grid Questionnaire	81
D. Leadership Adaptability and Style Inventory (LASI).....	82
E. The least preferred Co-worker.....	84
3. Les mesures multi-factorielles du leadership Transformationnel et Charismatique	86
A. Le Leadership Charismatique.....	86
1. C-R Scale of Charismatique Leadership.....	86
2. Ethical Leadership Scale (ELS).....	88
B. Le leadership Transformationnel.....	89
1. Multifactor Leadership Questionnaire (MLQ).....	89
2. Transformational leadership questionnaire (TLQ).....	92
3. Leadership Practice Inventory	93
4. Global Transactional Leadership Scale (GTL).....	95
5. Empowering Leadership Questionnaire.....	96
4. Les grands « audits » des qualités, compétences et personnalité des leaders.....	98
A. Les caractéristiques des leaders incompetents	98
B. Les grandes évaluations multi-trait multi-méthode.....	100
1. Profillor.....	100
2. Benchmarks.....	101
3. Le Tryptique.....	103
4. Le GELI (Global Executive Leadership Inventory).....	104

Chapitre 4. Une approche en plein essor : les 360 degrés.....114

1. De l'auto-évaluation à l'évaluation par les autres.....	114
A. L'auto-évaluation.....	114
B. L'Hétéro-évaluation.....	115
2. Le 360 : une présentation.....	118
A. La définition.....	118
B. Les avantages et les limites.....	119
1. Les sources d'évaluation.....	120
2. L'indulgence.....	121
3. L'effet de Halo	122
C. L'objectif du 360 : le changement	123
1. Facteurs favorisant le changement.....	124
2. Les effets du feedback sur la performance	124

3. Les formations.....	125
3. <i>Accord entre les évaluateurs</i>	127
A. L'accord au sein du groupe d'observateurs.....	127
B. L'auto-évaluation.....	128
C. L'accord entre les groupes d'observateurs.....	128
D. Variables pouvant influencer l'accord entre observateurs.....	130
1. Le genre.....	130
2. L'ethnicité.....	131
3. L'âge.....	131
4. L'expérience.....	132
5. L'éducation.....	132
6. La Similarité self - observateur	132
A. Le genre et la composition du groupe.....	132
B. L'âge et la composition du groupe.....	133
C. Nationalité et la composition du groupe.....	133
7. Variables de personnalité / aptitude.....	134
A. L'intelligence.....	134
B. La conscience de soi.....	134
C. La confiance en soi.....	134
D. Le locus de contrôle.....	135
E. Assertivité.....	135
F. Dominance	136
D. Les catégories d'observateurs.....	136
4. <i>Traitement des scores et évaluation des écarts</i>	137
A. Analyses multi-niveaux	137
B. Similarités inter- et intra-groupes.....	139
C. Théorie de la généralisabilité.....	140
D. Équivalence de la mesure.....	141

Méthode..... 144

1. <i>L'outil de mesure : le Global Executive Life Inventory (GELI)</i>	144
A. Les dimensions du GELI.....	144
B. Administration et passations.....	146
C. Qualités psychométriques.....	147

2. L'échantillon	148
A. Les critères d'inclusion dans l'échantillon.....	149
B. Description de l'échantillon.....	150
1. Age.....	150
3. Genre.....	151
4. Nationalités.....	152
5. Industries.....	152

Résultats..... 155

Chapitre 5. Les analyses multi-niveaux..... 157

1. Les analyses à un et à multiples niveaux.....	157
A. Les analyses ne tenant pas compte du multi-niveaux.....	158
B. Les analyses à multiples niveaux.....	159
2. Utilisation du multi-niveaux sur les échelles du GELI.....	160
3. Le multi-niveaux est-il nécessaire pour nos données ?.....	161
4. La variance Inter-selves et Inter-observateurs.....	163
5. L'importance des variables explicatives.....	166
A. La catégorie d'observateurs.....	166
B. Age.....	173
C. Genre.....	177
6. La variance encore libre	185

Chapitre 6. Étude de la consistance inter-observateurs..... 191

1. Les indices statistiques classiques de l'accord et de la consistance inter-juges..	191
2. Application aux résultats des questionnaires à 360°.....	194
A. Les corrélations Inter-Observateur.....	194
B. Les coefficients de corrélation Intraclasse.....	196
3. Une autre hypothèse pour comprendre l'accord entre les évaluateurs.....	198
A. L'hypothèse de la Restriction de la Variance.....	198
B. Les coefficients rWG.....	199
4. Calcul des coefficients rWG.....	201
A. La méthode LeBreton.....	201
B. La mise à 0 des valeurs « Hors-Norme ».....	204
C. Le choix de la distribution de référence pour estimer σ^2E	207

5. Catégorisation des groupes d'observateurs en tant que « Accord », « Désaccord » ou « Hors-Norme ».....	213
A. Résultat des analyses.....	213
B. Utilisation de l'échelle.....	215
6. Généralisation de la consistance entre observateurs.....	217
A. Calcul d'un indicateur général rWG(J) Total.....	217
B. Comparaison du groupement « Général » avec le groupement « Par échelles »	217
C. Stabilité du classement des groupes d'observateurs dans les catégories d'accord, pour les 10 dimensions du GELI.....	219

Chapitre 7 : Étude de la variable « Écart » – synthèse des différences inter-évaluateurs.....223

1. Les écarts inter-observateur dans leurs évaluations	223
2. Approche multi-niveaux des écarts entre observateurs au niveau des dimensions du GELI.....	225
A. La catégorie d'accord.....	227
2. Les variables explicatives classiques.....	231
1. Les catégories d'observateur.....	231
2. L'âge.....	234
3. Le genre.....	235
4. Le modèle complet.....	238
C. L'étude des écarts au niveau général.....	240
1. Le modèle général	241
2. La fréquence d'utilisation de l'ancrage 2	243

Chapitre 8 : Le score vrai et la théorie de la mesure.....246

1. Le score vrai et la théorie de la mesure.....	247
A. La théorie classique	247
B. L'estimation ponctuelle du score vrai.....	250
1. Les biais dans l'utilisation des échelles de Lickert – les ancrages.....	251
2. La technique des Vignettes d'ancrage.....	252
C. Le modèle du consensus culturel.....	256
2. L'estimation du score vrai sous forme de distribution.....	258

A. La valeur manquante fondamentale de Rubin.....	258
B. Analyse Factorielle Confirmatoire multi-niveaux.....	261
1. Le modèle.....	262
2. L'adéquation du modèle.....	267
<u>Discussion générale.....</u>	<u>272</u>
- Les limites de notre étude.....	277
<u>Index des Tableaux et Graphiques.....</u>	<u>283</u>
<u>Index des Équations.....</u>	<u>288</u>
<u>Index des Auteurs.....</u>	<u>293</u>
<u>Références.....</u>	<u>301</u>
<u>Annexe.....</u>	<u>335</u>

Introduction

Que ce soit dans les essais sur l'histoire, sur l'économie, la politique – puis la sociologie et enfin la psychologie du travail et des organisations – la notion de leader a toujours été présente. Qu'il soit défini par son pouvoir ou son action il s'agit toujours d'un homme (parfois d'une femme) se positionnant face à d'autres ou des groupes d'autres, et créant par son action ou ses idées un mouvement au sein de ce groupe.

On peut noter que le mot leadership est issu de l'anglais *to lead* issu lui même du latin *laed*, signifiant chemin. Un leader serait donc étymologiquement celui qui montre à ses pairs le chemin en marchant en tête (Kets de Vries, Vrignaud, & Florent-Treacy, 2004).

Dans le cadre de ce travail de thèse nous allons passer en revue de nombreux éléments sur le leadership et son évaluation. En premier lieu, nous allons nous poser la question de la définition du leadership. Qui est le leader ? Est-ce notre personnalité qui fait de nous un leader ? Est-ce la situation dans laquelle nous nous trouvons qui nous rend leader ? Est-ce notre style ou nos comportements ? Ou bien est-ce un ensemble de tout cela ?

Nous verrons donc comment les différents modèles du leadership sont passés d'une conception unitaire, à une conception interactionniste puis à une définition en termes d'effet du leadership.

Mais le leadership est-il réellement un concept indépendant des autres ? Dans une approche plus différentielle, nous allons nous pencher sur les liens et les inter-relations que le leadership entretient avec d'autres variables psychologiques comme la personnalité ou l'intelligence émotionnelle, pour enfin nous poser la question de son efficacité. En effet, la majorité des théories se proposent de définir non seulement un leader, mais surtout un leader efficace. Alors, qu'est-ce qu'un leader efficace ? Comment peut-on définir un leader efficace ? Comment peut-on mesurer son efficacité ?

La question de la mesure sera alors abordée plus directement. En effet, dans le cadre de leurs théories sur le leadership, une majorité de chercheurs ont utilisé des outils de mesure, pour éprouver la validité de leurs modèles. Ces outils, avec le temps et la popularité de certaines théories, sont devenus des questionnaires psychométriques à part entière et sont aujourd'hui utilisés pour évaluer les qualités de leadership des managers.

Mais un manager peut-il juger par lui-même de son potentiel de leadership ou même de son leadership ? La question s'est posée pour de nombreux chercheurs lors de la validation de leurs modèles : n'est-ce pas les autres qui nous rendent leader ? Pourtant ce n'est que des années plus tard qu'une méthode d'évaluation spécifique tenant compte de ce constat a pu être proposée : l'évaluation dite à 360 degrés. Un des points majeurs de notre recherche sera d'étudier les apports et les limites de cette méthode à partir de ses origines, des différentes approches utilisées et de sa confrontation aux modèles psychométriques et statistiques les plus récents.

En effet, l'approche à 360° se propose d'effectuer une évaluation des qualités d'un manager en interrogeant un ensemble de personnes travaillant avec lui (subordonnés, collègues, supérieurs...) et en comparant leurs évaluations avec la propre évaluation du manager. Or c'est à cette étape que les difficultés de cette méthode apparaissent.

Nous verrons ainsi que le point d'intérêt central de cette approche réside dans l'étude des différences entre les catégories d'observateurs et entre chacune de ces catégories et l'auto-évaluation du manager. Pourtant cette démarche même n'est pas usuelle en psychométrie du fait de la structure des données issues des analyses à 360° où les évaluations des observateurs sont imbriquées dans l'auto-évaluation du manager.

Nous verrons donc dans quelle mesure les variables souvent utilisées pour expliquer ces désaccords entre catégories d'observateurs permettent réellement de comprendre les écarts observés. Plus spécifiquement, nous étudierons dans quelle mesure ces variables permettent (en utilisant les méthodes multi-niveaux) de réduire (expliquer) la part de variance inter-self (entre les auto-évaluations des managers) et la part de variance inter-observateur / intra-self (entre les évaluations des observateurs d'un même manager).

Ainsi, bien que de nombreuses hypothèses quant à l'effet des variables comme le genre, l'âge ou le type d'observateur ont été vérifiées par les recherches précédentes, nous verrons dans le cadre de cette thèse, qu'au final, ces variables n'expliquent que très partiellement la variance. Ceci nous conduira à mettre l'accent non plus sur les différences inter-groupes mais sur le concept d'accord et de désaccord au sein des groupes d'observateurs.

En effet, pour comparer les évaluations faites par les observateurs avec celle faite par le manager, faut-il encore pouvoir démontrer un accord au sein du groupe d'observateurs. Nous allons donc nous attacher à mesurer l'accord au sein des groupes d'observateurs, utilisant aussi bien des méthodes classiques d'accord inter-juges (corrélation Bravais-Pearson, corrélation Intraclasse), que des approches plus récentes et non corrélationnelles (r_{wg}). Nous aborderons à cet effet la controverse entourant l'accord intra-groupe en discutant l'hypothèse de LeBreton (2003) qui stipule qu'il n'y a pas, dans le cadre des 360°, de désaccords réel, mais que le désaccord trouvé par les chercheurs est dû à un artefact statistique expliqué par l'hypothèse de la « réduction de la variance ».

Cette approche de LeBreton, dont nous allons également mettre en lumière les limites, va tout de même nous permettre d'isoler trois catégories de groupes d'observateurs. Un groupe qui est en Accord, un qui est en Désaccord quant à son évaluation du manager et enfin un troisième groupe Hors-norme dont nous aborderons les spécificités.

Nous avons choisi dans la dernière partie de prendre de la hauteur et de nous poser une question plus méthodologique. Quel est donc le modèle psychométrique à la base des évaluations en 360° ? Comment définir le score vrai dans le cadre de ces évaluations ? Peut-on parler de variance vraie ? Comment la théorie classique du score vrai et de l'erreur de mesure peut-elle s'appliquer aux données de questionnaires à 360° ?

Ainsi, après avoir rappelé le modèle psychométrique classique du score vrai et de l'erreur de mesure, nous chercherons à le mettre en relation avec le modèle des questionnaires à 360. Nous proposerons donc plusieurs approches pour estimer le score vrai dans le cadre du 360. Une approche visant l'estimation ponctuelle du score vrai, en phase avec la théorie classique de la mesure, et une autre approche plus innovante où l'estimation du score vrai sera réalisée sous forme de distribution. Pour cette dernière, nous ferons appel à une méthodologie encore peu utilisée dans les recherches sur les 360° : les analyses factorielles confirmatoires multi-niveaux. Celles-ci nous permettront de proposer une définition psychométrique des scores issus des analyses à 360° intégrant la spécificité de cette méthode au modèle psychométrique.

Chapitre 1. Approche historique du concept de leadership

À son début, une part des recherches et des études sur le leadership se sont clairement positionnées dans le domaine de la psychologie sociale. La notion de leader se mélange alors à la notion de chef et s'organise, d'une part, autour des notions de chef légitime et formel opposé au chef informel et, d'autre part, de la dynamique des groupes et de la notion d'influence.

Parallèlement l'intérêt croissant du grand public et du milieu des affaires pour les grands leaders et les hommes d'exception a orienté l'étude du leadership vers l'approche biographique (Cameron, 1936). Ainsi, beaucoup ont tenté de percer les mystères des grands de ce monde en se penchant sur leur parcours aussi bien personnel que professionnel. Ces démarches ont notamment débouché, en psychologie, sur les premières conceptions du leader en termes de traits et de personnalité et sur les théories des « Grands Hommes » (Stogdill, 1974).

C'est vers le 20ème siècle que la psychologie du travail, toute jeune, a intégré ce concept. La psychologie du travail et des organisations, s'est fixée pour objectif « l'étude des conduites humaines et de production d'un bien ou de mise en œuvre d'un service, développés dans le cadre d'une organisation marchande ou non marchande » (Lelouche, 2007) cette notion de conduite englobant d'une part les aspects observables aussi bien que les éléments plus internes (cognitifs et conatifs) qui les accompagnent. Parmi les trois champs d'étude de la psychologie du travail et de l'organisation, tel que définis par l'ENOP (*European Network of Organizational and Work Psychologists*, (1998) : la relation homme-tâche, la psychologie du personnel et la psychologie organisationnelle, c'est dans ce dernier champ que le leadership se place. Il se situe dans le cadre de l'étude des comportements collectifs et de leurs développements, aux côtés de concepts comme la culture d'entreprise, le fonctionnement des équipes ou le changement organisationnel.

Pourtant, bien que le concept de leadership soit à présent durablement encré en psychologie du travail, son héritage sociologique n'est pas loin et se retrouve dans de nombreuses définitions de celui-ci. En effet, si nous examinons les diverses définitions du

leadership, nous retrouvons souvent des aspects et des concepts de psychologie sociale tout autant que de psychologie du travail. Déjà en 1974, Stogdill dans un des premiers traités sur le Leadership recensait plus de trois cent cinquante définitions de ce qu'est le leadership et proposait finalement de définir ce dernier comme « le fait d'initier et de maintenir la structure tout en interagissant vers un but » (Stogdill, 1974, p. 114). Nous pouvons aussi définir le leadership comme « le comportement d'un individu qui cherche à orienter les activités d'un groupe d'individus vers un objectif commun » (Hemphill & Coons, 1957, p. 7), ou bien comme l'influence interpersonnelle exercée dans une situation donnée et dirigée (à travers les voies de communication) vers l'atteinte d'un objectif (Tannenbaum, Weschler, & Massarik, 1961). De même, pour Rauch et Behling (1984, p. 46) le leadership peut être défini comme « le fait d'influencer les activités d'un groupe organisé en vue de l'atteinte d'un objectif donné ». Pour Burns (1978) le leadership est quant à lui exercé lorsque des individus mobilisent l'ensemble des ressources disponibles (institutionnelles, politiques, psychologiques) pour engager leurs subordonnés dans l'action et atteindre leur objectif. De même pour House (1971) le leadership est une aptitude individuelle à influencer, motiver et amener les autres à participer au succès et à l'efficacité d'une organisation.

Un autre ensemble de définitions donnent, quant à elles, une place plus importante au sens. Ainsi, Jacobs et Jacques (1991) définissent le leadership comme le processus permettant de donner un but (qui ait du sens) à l'effort collectif. Le leadership est aussi décrit comme le processus qui permet de donner un sens à ce qu'accomplissent les individus ensemble (Daft & Paulus, 1994), et comme la mise au jour d'une vision, sous l'égide de laquelle les choses peuvent être accomplies (Richards & Engle, 1986).

Ainsi, englobant ces diverses conceptions, nous allons nous attacher à la conception clinique de Manfred Kets de Vries (2003b), qui définit le leader à travers deux rôles principaux : un rôle de chef (rôle charismatique) et un autre d'architecte (rôle architectural). Dans son rôle de chef, le leader donne des perspectives pour l'avenir, partage et délègue le pouvoir et parvient à mobiliser ses subordonnés. Dans son second rôle, le leader privilégie les questions liées au projet de l'entreprise ou de l'organisation. Il s'occupe également de la restructuration des systèmes de récompense et de contrôle.

En effet, cette définition résume les points marquants vus précédemment. A savoir que le leadership est un processus où s'exercent des relations d'influence, dans un contexte de groupes et ce avec des individus qui partagent des objectifs communs et qui parviennent à se mobiliser pour réaliser les objectifs souhaités (Inspiré de Daft, 2005; Northouse, 2004; Yukl, 2002).

Nous allons donc dans ce premier chapitre nous intéresser aux différentes approches qui ont donné lieu à ces conceptions et ces définitions variées du leadership. Nous nous intéresserons aux théories générales sur les comportements des leaders. Nous commencerons par les approches personnalistes, situationnelles, interactionnistes puis transformationnelles et transactionnelles du comportement du leader, pour terminer sur une conception plus clinique et intégratrice du concept. Adoptant une démarche différentielle, nous chercherons à prédire les comportements du leader mais aussi à expliquer son efficacité à travers de nombreuses variables d'effet. Pour ce faire nous aborderons l'ancrage du leadership dans la personnalité, et ses liens avec l'intelligence émotionnelle.

Poursuivant notre démarche, nous aborderons ensuite la question de de la mesure. Outil premier de l'approche du leadership, nous verrons comment les questionnaires psychométriques se proposent de mesurer, en fonction de leurs différents ancrages théoriques, les qualités des managers. Nous nous apercevrons ainsi que de nombreux chercheurs préfèrent mesurer le leadership non pas à travers l'évaluation qu'en fait le manager lui-même, mais à travers les évaluations qu'en font ceux qui le côtoient ou qui en sont subordonnés. Enfin, dans la quatrième partie nous nous plongerons plus encore dans ces méthodes d'évaluation spécifiques et de plus en plus populaires que sont les évaluations à 360° (lire 360 degrés), où tout l'entourage du manager tout autant que le manager lui même, évaluent ses qualités de leader. Nous y verrons les intérêts mais aussi les particularités et les difficultés de cette méthodologie.

1. Les théories unitaires

Nous allons dans un premier temps nous pencher sur les théories unitaires du leadership. Qu'elles placent en position centrale la personnalité, la situation et l'environnement extérieur ou le style du leader, pour ces théories un de ces éléments est suffisant pour définir et comprendre l'efficacité des leaders.

A. Les traits de personnalité

1. *Great Man Theory*

Suite aux travaux sur la génétique de Galton, vers la fin du 19^{ème} siècle, beaucoup de théoriciens ont tenté d'expliquer le leadership grâce à une approche héréditaire.

Selon cette approche, le leader serait porteur de qualités spécifiques qui lui seraient propres. De plus, pour un certain nombre de théoriciens, ces qualités seraient héréditaires – c'est l'idée des « leaders naturels ».

Cette démarche a donné naissance à l'approche personnaliste du leadership. Approche qui tente d'expliquer le leadership en termes de traits de personnalité et de caractère, et qui se veut indépendante du contexte. Quelque soit l'environnement, certains individus seront des leaders et seront efficaces, alors que d'autres non, d'où cette notion de « leader naturel ».

Vers la fin des années 40, lors de sa première revue de la littérature, Stogdill (1948) met en avant un ensemble de caractéristiques issues des recherches de l'époque, qui seraient propres au leader. Ces caractéristiques regroupent aussi bien des traits de personnalité (sociabilité, persistance..), que des capacités cognitives (Northouse, 2004), mais l'approche par traits ayant connu des résultats mitigés et ne permettant plus de suivre les définitions de plus en plus complexes du leadership, à été progressivement abandonnée.

Cependant, sous une forme quelque peu différente l'approche du leadership par la personnalité du leader a connu un regain d'intérêt.

2. Le leadership comme un construit secondaire des tests de personnalité

Les liens entre la personnalité et le leadership ne se limitent pas aux théories des leaders naturels. Un certain nombre de questionnaires de personnalité proposent par exemple de « déduire » les caractéristiques ou le potentiel de leadership d'un individu. Ils cherchent donc à l'estimer à partir de scores de celui-ci aux dimensions de personnalité. Il s'agit donc d'une mesure « indirecte » du leadership qui se distingue de celles qui seront présentées dans le chapitre 3.

Ainsi, dans le 16PF (16 *Personality Factor*, Cattell, Eber, et Tatsuoka, 1970) qui est un grand questionnaire de personnalité permettant d'estimer les scores des sujets sur 16 dimensions de la personnalité (ce questionnaire sera présenté plus en détail dans le chapitre deux (Chapitre 2.1.A.1), les auteurs proposent de calculer, *a posteriori*, un score secondaire de leadership (Cattell et al., 1970).

Dès la création du 16PF, Cattell et Stice (1954) ont développé une équation permettant de calculer un potentiel de leadership sur la base des dimensions de personnalité mesurées par le test. L'étude originale avait été organisée dans le cadre militaire. Lors de celle-ci, 800 militaires avaient été répartis en dix groupes et devaient, en groupe, réaliser des tâches de prise de décision. Chaque groupe avait également élu un leader durant les premières phases de l'étude, mais tous les leaders élus n'ont pas été maintenus au fil des exercices et parmi les leaders maintenus, tous n'avaient pas été efficaces.

Les leaders les plus efficaces étaient caractérisés par les dimensions suivantes : Stabilité Émotionnelle, Dominance, Vivacité, Respect des règles, Aplomb en société, Sensibilité (-), Abstraction (-), Appréhension (-) et Perfectionnisme (Cattell et al., 1970). Par la suite, d'autres études ont validé ce construit secondaire et la sixième version de ce test (*16PF fifth version* ; Cattell, Cattell, et Cattell, 1993) le propose toujours.

Enfin, même si les tenants du domaine ne conçoivent plus vraiment le leader sur la seule base de ces qualités et traits de personnalités, la recherche actuelle reste très prolifique sur l'étude des liens entre la personnalité et le leadership comme nous le verrons dans le second chapitre.

B. Les théories Environnementales

Une autre approche définit le leader comme totalement issu de la situation. Une grande partie des auteurs issus de ce courant considèrent que l'émergence d'un leader est le résultat d'une combinaison entre une situation (un temps donné, un ensemble de problèmes, des conflits à résoudre) et des caractéristiques personnelles.

Pearson (1928) propose dans sa théorie deux hypothèses qui, d'après lui, permettent de comprendre l'émergence du leadership : d'une part, toute situation particulière conditionne les qualités nécessaires au leader et de ce fait le type de leader qui émergera par la suite. D'autre part, ces qualités individuelles que la situation définit comme qualités d'un leader sont elles-mêmes conditionnées par les situations de leadership précédentes qui les ont modelées et développées.

Par ailleurs, d'après Murphy (1941) le leadership n'est pas une caractéristique de la personne mais est une fonction de la situation. En effet, chaque situation appellerait un certain type de comportements et d'actions, le leader n'injecterait donc pas de leadership au sens propre mais serait le facteur instrumental qui permet d'atteindre une solution.

Dès les années 50, de nombreux auteurs ont remis en cause l'approche environmentaliste. En effet, la situation seule ne serait pour eux pas un élément suffisant pour expliquer l'émergence de leaders.

C. Les approches comportementales : les comportements et les styles du leader

Nous avons vu jusqu'à présent deux conceptions opposées du leadership : l'une centrée sur les dimensions de personnalité du leader et l'autre centrée sur les caractéristiques de la situation. Dans les deux cas, ces approches visent à prédire l'efficacité d'un manager donné en fonction de ces éléments seulement. Une troisième approche unitaire consiste à considérer que le leader se caractérise par son comportement, son style de leadership et que celui-ci est le seul déterminant de son échec ou de son succès.

Nous verrons dans la partie suivante que de nombreuses approches ont été proposées pour décrire les styles des leaders. Mais malgré leurs différences, dans tous les cas, la question principale de ces approches est de trouver le style le plus efficace. Pour mettre en lien ces styles avec la notion d'efficacité, nous allons évoquer pour chacune des théories quelques exemples de variables d'effets : dimensions qui sont considérées comme des effets positifs du leadership, et qui seront abordées en détail dans le chapitre deux (Chapitre 2.2).

1. Les études de l'Université d'Iowa

A. Lewin, Lippitt et White (1939)

Déjà dès la fin des années 1930, à l'Université de l'Iowa, Lewin, Lippitt et White (1939) proposent une étude expérimentale des styles de leadership.

Des jeunes d'un club de modélisme sont répartis selon trois conditions expérimentales : un groupe est encadré par un adulte se comportant de manière autocratique, un second groupe par un adulte avec un comportement démocratique et le troisième groupe par un adulte au comportement « laissez-faire ».

Dans les cas des adultes autocratiques et laissez-faire, les garçons ont montré des réactions agressives les uns avec les autres, contrairement au groupe démocratique. En ce qui concerne la quantité de travail réalisé, les groupes autocratique et démocratique ont présenté des résultats similaires.

Ainsi, le leadership laissez-faire n'étant pas à proprement parlé un style de leadership, les auteurs de l'Iowa ont proposé une compréhension du concept de leadership reposant sur deux styles de base :

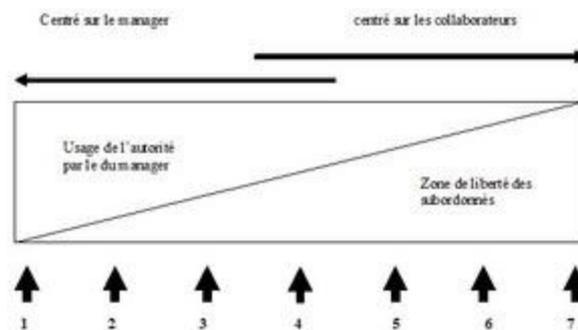
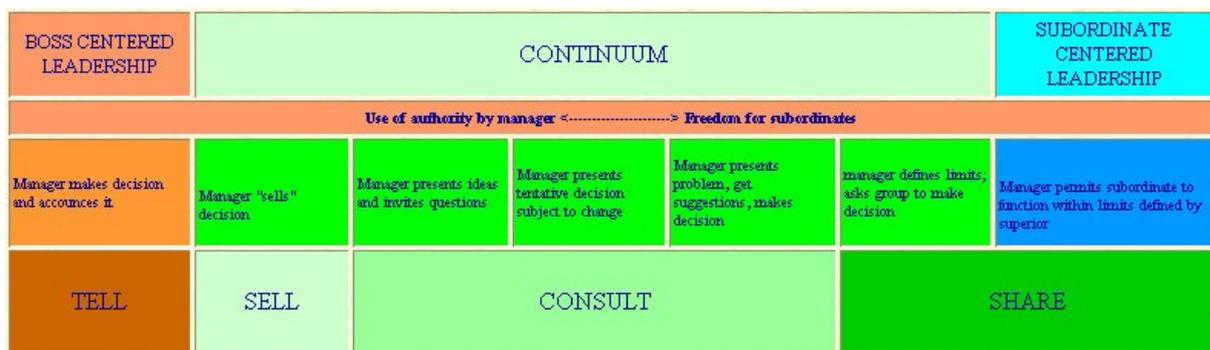
D'une part il y a le **leader autoritaire** aussi appelé le leader centré sur la tâche. Il prend seul les décisions et supervise étroitement le travail de ses subordonnés, il assume l'ensemble des responsabilités pour toutes les étapes du processus.

D'autre part, il y a le **leader démocratique** aussi appelé le leader centré sur les personnes. Il encourage la collaboration, l'organisation du travail par les subordonnés et la

participation dans les décisions.

B. Le continuum de Tannenbaum et Schmidt (1958)

Basé sur les premiers travaux de Lewin et de ses collaborateurs (1939) le continuum de Tannenbaum et Schmidt (Graphique T.1) propose un ensemble de styles intermédiaires (sept au total) entre le leadership autoritaire et démocratique.



Graphique T. 1 : Styles de leadership de Tannenbaum et Schmidt

Tannenbaum, R., Schmidt WH (1958), How to choose a leadership pattern. Harvard Business Review 36 p.100

Dans le premier style le leader décide seul, dans le second style le leader « vend » ses décisions à ces subordonnés. Dans le style trois, le leader présente ces idées et essaye de susciter des questions. Le style quatre se caractérise par des prises de décisions présentées comme non définitives et modifiables en fonction des avis des subordonnés. Avec le style cinq, le leader expose la situation et demande expressément des suggestions avant de prendre sa décision. Enfin, les styles six et sept sont appelés « styles de partage ». En appliquant le style six, le leader ne fait plus que définir un cadre et laisse ces subordonnés la charge de la prise de décision. Le style sept, enfin permet

aux collaborateurs d'agir dans le cadre réservé au supérieur hiérarchique.

De nombreuses études ont été menées pour éprouver ce modèle, et bien que l'idée de style reste particulièrement utilisée, les conclusions sur le style « le plus efficace » sont plus contrastées.

À partir des résultats de son étude des styles de leadership dans une entreprise d'informatique Anglaise, Sadler (1970) a montré que la plupart des employés ont pu classer leur leader dans l'une des quatre catégories (*tells, sells, consults, joints*). Il s'est par ailleurs intéressé au style préférentiel (différent du style observé) et a rapporté des résultats très différents. Ainsi, le style préférentiel est le style Consultatif (*consults*) et ce quelle que soit la position hiérarchique du répondant (ouvrier, employé ou cadre). Par contre, c'est le style Démocratique (*joints*) qui est le moins apprécié de tous, ce qui est un résultat étonnant.

Ainsi, la satisfaction des employés avec leur travail et leur entreprise est similaire quelque soit le style le leadership, mais elle est significativement plus faible dans le cas où le leader ne présente aucun style reconnaissable.

De même que Tannenbaum et Schmidt (1958), des styles similaires ont pu être définis dans la littérature plus « managériale ».

2. Théorie X et Théorie Y

Partant du postulat qu'il existe deux types de population (ceux qui veulent mener et prendre des responsabilités et ceux qui veulent être dirigés et ne veulent pas prendre de responsabilités) McGregor (1960) a fait une classification du leadership en deux styles : autoritaire (Théorie X) ou égalitaire (Théorie Y).

La théorie X, présuppose que l'être humain est paresseux, évitant et non désireux de travailler. Partant de là, les leaders se doivent d'appliquer des mesures de contrôle et de gestion strictes et directives envers leurs subordonnés. D'après la théorie X, les employés sont exclus des décisions et toutes les responsabilités restent du ressort du leader.

La théorie Y, quant à elle, est basée sur le concept de « réalisation de soi » (*self-actualisation*) de Maslow (1943) qui considère que le travail peut être apprécié, que les gens sont prêts à travailler dur et à assumer des responsabilités s'ils ont la possibilité d'atteindre leurs buts personnels en même temps qu'ils atteignent les objectifs de l'entreprise. D'après la théorie Y, un leader se doit donc de créer un environnement favorable à la réalisation personnelle de chacun en synergie avec la réalisation des objectifs de l'entreprise. Il est donc centré sur les besoins des personnes, les responsabilités sont partagées et les décisions prises de manière plus participative.

3. Les études de l'Université du Michigan

Likert (1961, 1967) et les chercheurs de l'Université du Michigan s'intéressent aux liens entre les comportements des leaders et l'efficacité du groupe. Elles visent plus spécifiquement à identifier les styles de leadership associés à la performance et à la satisfaction.

Les études ont été menées auprès de nombreuses organisations et entreprises afin d'isoler les comportements clefs (Katz et Kahn, 1952). Deux types de comportements ont ainsi été mis en évidence :

D'une part les **comportements centrés sur la tâche** : les leaders n'effectuent pas les mêmes tâches que leurs subordonnés. Ils s'occupent principalement de l'organisation, de la planification et de l'apport des ressources nécessaires à la réalisation du travail.

D'autre part, les **comportements centrés sur la personne** : le leader se préoccupe du relationnel entre et avec ses subordonnés, il encourage la participation dans l'organisation et la prise de décision.

Présentés dans un premier temps comme les extrêmes d'un même continuum, ces dimensions sont devenues au fil des études des dimensions distinctes et non exclusives.

Les travaux de Likert (1961, 1967) sur le leadership participatif, sont certainement parmi les contributions les plus connues de l'Université du Michigan.

D'après Likert, la performance maximale est atteinte lorsque l'attention est portée sur la personne. Quatre styles de leadership ont pu être définis sur la base des deux types de comportements – centrés sur la tâche et sur la personne.

Lorsque le comportement est principalement centré sur la tâche, le style du leader peut être :

Exploiteur-Autoritaire – ce style est manipulateur par la peur et le contrôle. La communication et les décisions partent de la hiérarchie vers la base et il résulte en une productivité faible.

Le second style centré sur la tâche est qualifié de **Paternaliste**. Il s'agit d'un management « par la carotte ». C'est une version un peu allégée du style Exploiteur. La communication est toujours descendante, et seules quelques décisions mineures sont déléguées. Globalement, la productivité est moyenne.

D'autre part, les comportements centrés sur la personne peuvent être répartis, eux aussi, en deux styles de leadership : le style **Consultatif** et le style **Participatif**. Dans le style **Consultatif** les salariés sont consultés par la hiérarchie. La communication se fait dans les deux sens – même si les décisions importantes se prennent tout de même de manière centralisée.

Enfin, au niveau **Participatif**, les méthodes participatives sont utilisées au maximum, les salariés participent aux décisions à travers différents groupes. La communication se fait dans les deux sens et les membres de l'organisation se sentent appartenir au groupe et se considèrent bien et satisfaits au niveau professionnel.

C'est donc le style participatif qui est, d'après Likert, le style de leadership optimal.

4. Les études de l'Université de l'Ohio

Dans le même temps, les chercheurs de l'Université de l'Ohio mènent des recherches pour identifier, eux aussi, les dimensions indépendantes (orthogonales) du comportement des leaders et déterminer leurs effets sur la performance et la satisfaction des employés (Halpin et Winer, 1957; Stogdill, 1974, 1950).

Ces chercheurs sont partis d'une liste de plus de 1800 phrases descriptives qui ont dû être classées en catégories par des juges. Puis, les dix dimensions générales obtenues ont pu être réduites à deux dimensions comportementales : les comportements d'Initiation de structure d'une part, et de Considération de l'autre.

Les comportements **d'initiation de structure** décrivent dans quelle mesure le leader est orienté vers la tâche et dirige le travail de ses subordonnés dans ce sens. Les comportements de **considération** décrivent quant à eux, dans quelle mesure le leader se « préoccupe » de ses subordonnés, de leurs émotions et de leurs idées.

Le fait que les deux comportements ne soient pas mutuellement exclusifs constitue la différence entre les travaux des chercheurs de l'Ohio et les deux premières études présentées. En effet, avec le modèle de l'Ohio, un leader peut maîtriser ces deux types de comportement.

Le modèle de l'Ohio commence également à tenir compte de la situation. En effet, les chercheurs concluent que « certaines situations pourraient exiger que l'on mette davantage d'accent sur l'un ou l'autre de ces comportements » (Schriesheim et Murphy, 1976).

5. Les travaux de Blake et Mouton

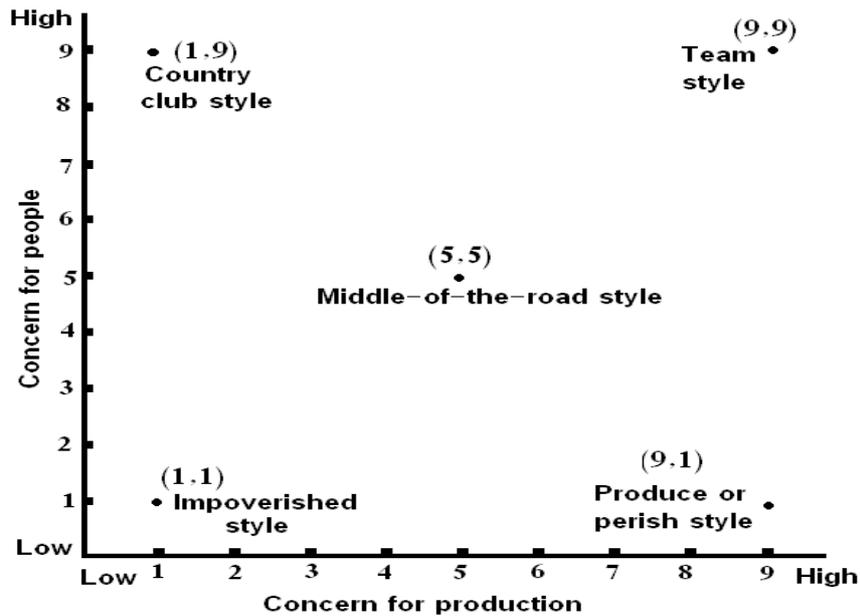
Vers la fin des années 60, Blake et Mouton (1961) proposent une théorie du leadership incorporant les résultats des travaux de l'Université de l'Ohio et de l'université du Michigan.

Blake et Mouton (1978), proposent une grille managériale croisant la préoccupation pour les individus avec la préoccupation pour le produit.

Dans leur grille (Graphique T.2), il existe une indépendance entre les deux dimensions et les auteurs attribuent un score de zéro à neuf à chacune (sur la base d'un questionnaire), ce qui leur permet de décrire cinq types de leaders en fonction des scores respectifs sur chacun des deux facteurs.

- **Le style appauvri** (*Impoverished style* – 1.1) se définit par des efforts minimum pour atteindre les objectifs. Le manager laisse faire les salariés, et ne réalise que des contrôles minimum.
- Dans **le style autoritaire** (*Perish style* – 9.1) le leader ne s'intéresse qu'aux résultats. Il impose son autorité, et contrôle tout.
- **Le style country-club** (*Country Club style* – 1.9) se caractérise par un leader préoccupé davantage de l'ambiance de travail que des résultats, il cherche surtout à plaire au groupe.
- **Le style institutionnel** (*Middle of the road style* – 5.5) décrit un leader qui recherche un équilibre entre les performances et le climat social.
- Enfin dans **le style équipe** (*Team style* – 9.9) : le leader emploie une approche intégratrice qui cherche à créer une bonne ambiance de travail, et à impliquer les salariés pour atteindre les objectifs.

C'est ce dernier qui est considéré comme idéal et le plus efficace.



Graphique T. 2 Styles de leader de Blake et Mouton

Blake, R.R., Mouton, J.S. (1961). *Group dynamics - Key to decision making*, Houston: Gulf Publishing Co.

6. Les limites et critiques de l'approche du leadership sous forme de style

De nombreuses critiques ont pu être apportées aux approches comportementales s'intéressant aux styles de leadership. L'une des premières critiques a été présentée par les tenants mêmes de l'approche issus de l'équipe de l'Université de l'Ohio (Schriesheim et Murphy, 1976) et porte sur l'absence d'analyses situationnelles.

Par la suite, beaucoup de chercheurs ont reproché aux théories comportementales de ne pas suffisamment tenir compte de variables pouvant modérer les liens entre les styles de leadership et les variables d'efficacité (satisfaction des employés ou performance de l'équipe).

Bryman (1992), dans les années 1990, a présenté un certain nombre d'autres limites de ces approches comportementales : les problèmes de causalité, les problèmes des groupes, l'absence d'analyse situationnelle (que nous avons déjà abordé) et enfin l'absence de prise en compte du leadership informel.

En ce qui concerne la causalité, les études mettant en relation les styles de leadership et la satisfaction par exemple, prennent la plupart du temps ces mesures de manière concomitantes. Aussi, le postulat selon lequel ces variables sont des effets (*outcomes*) des différents styles de leadership repose sur un ensemble de corrélations dont la causalité est déduite après coup (Bryman, 1992; Shackleton, 1995). Ainsi peu d'études ont adopté une réelle approche longitudinale afin de mettre à jour les liens de causalité entre les styles et les effets du leadership. Ces études (Greene, 1975) ont par ailleurs donné des résultats mitigés, validant certaines relations causales (par exemple les leaders orientés vers la tâche créent une satisfaction plus importante chez les employés) mais en remettant en cause d'autres – ainsi la faible performance des employés a pu amener le leader à adopter des comportements plus centrés sur la tâche et non inversement.

De même, la grande majorité des recherches se base sur la prise en compte des leaders « officiels » lors des études et des comparaisons des différents groupes. Pourtant, il existe souvent des leaders informels dans les groupes professionnels aussi bien que dans les groupes informels. Il serait donc envisageable que l'influence exercée par ce leader informel soit d'un style différent de celle du leader hiérarchique / officiel. Ainsi, les conclusions tirées sur les effets de tel ou tel style de leadership (basé sur l'étude des caractéristiques du leader formel) seraient à remettre en question si le style du leader informel se trouve être différent.

Enfin, en ce qui à trait aux groupes, la plupart des études s'intéressent aux liens entre le groupe et le leader. Mais, lorsqu'elles abordent ce lien, ces études tendent à moyenniser les évaluations par groupes pour obtenir le style du leader. Or, comme l'ont montré Katerberg et Hom (1981) la variabilité des réponses pourrait être importante et les conclusions tirées différentes si les réponses individuelles (non moyennées) étaient prises en compte. Cette limite méthodologique, qui se retrouve dans beaucoup d'hétéro-évaluations, sera plus longuement abordée lors de l'étude des méthodes d'évaluation à 360°.

2. L'intégration des variables situationnelles et de contingence

Au-delà des styles propres aux leaders, la plupart des recherches qui ont fait suite aux études de l'Ohio et du Michigan, intègrent les situations dans lesquelles le processus de leadership s'accomplit. Ainsi, dès les années 30, Westburgh (1931) considère qu'il est nécessaire de prendre en compte, d'une part, les traits individuels (au niveau affectif, intellectuel et de l'action) et, d'autre part, les conditions spécifiques dans lesquelles cet individu opère.

De même, Case (1933) toujours dans les années 30, définit trois facteurs qui, d'après lui, permettent de produire une conjoncture favorable à l'émergence d'un leader : les traits de personnalité du leader, la nature du groupe ainsi que de ses membres et, enfin, les événements auxquels est confronté le groupe.

Dans le cadre de ces recherches le leader n'est donc plus considéré dans l'absolu, car un leader ne l'est pas nécessairement dans toutes les situations. Ce qui importe avec les approches de contingence et situationnelles, c'est l'adéquation entre le comportement du leader, son style et le type de situation.

Avec l'approche situationnelle, c'est la situation qui détermine le style de leadership à adopter. Avec l'approche de la contingence, les variables sont inter-reliées et la direction entre les liens est indéterminée (Ayman, 2004). Le leadership représenterait donc la maîtrise de l'interaction entre les objectifs propres du leader et ceux des suiveurs.

A. La théorie de Contingence de Fiedler (1967)

Fiedler est l'un des premiers à proposer – suite aux études sur les styles de leadership – un modèle tenant compte de variables situationnelles.

En effet, le Modèle de Contingence de l'efficacité du Leader développé par Fiedler repose sur l'idée que l'efficacité d'un groupe est fonction de l'interaction entre le style du leader et la mesure dans laquelle la situation permet au leader d'exercer une influence sur le groupe. Deux variables sont donc étudiées dans la théorie de Fiedler : d'une part le style de leadership et d'autre part les caractéristiques de la situation (Ashour, 1973).

Fiedler identifie le style du leader (orientation vers la tâche ou vers la relation) grâce à un outil bâti autour de 16 adjectifs en opposition, permettant de déterminer le style préférentiel d'un leader : le « *Least Preferred Co-worker* » (Fiedler, 1967) - celui-ci sera abordé plus en détail dans la partie sur les instruments d'évaluation).

Les situations sont, quant à elles, évaluées en fonction de leur caractère favorable (*favourableness*) donnant la possibilité au leader d'exercer du contrôle et de faire jouer son influence.

Pour caractériser une situation trois variables sont étudiées : la relations entre le leader et les membres du groupe en lien avec la confiance que chacun accorde à l'autre, la structure des tâches et enfin le pouvoir hiérarchique lié à la position même du leader.

Ces trois variables permettent de définir huit types de situations (Fiedler, 1967) allant de la « plus favorable » à la « moins favorable ». La situation la plus favorable est caractérisée par une bonne relation de confiance entre le membre et le leader, une tâche fortement structurée et une forte position de pouvoir. Inversement, la situation la moins favorable se caractérise par une relation peu confiante, une tâche faiblement structurée et une position du leader associée à peu de pouvoir.

Entre ces deux extrêmes, le caractère favorable des situations s'ordonne en premier en fonction de la relation de confiance entre les membres et le leader (riche ou pauvre), puis en fonction de la structure de la tâche (structurée / non structurée) et enfin en fonction de la position hiérarchique de pouvoir du leader (forte position de pouvoir / faible position de pouvoir)

D'après Fiedler (1978), les leaders orientés vers la tâche réussiront mieux que les leaders orientés vers les relations humaines, lorsque la situation leur sera très favorable ou très défavorable. À l'inverse, les leaders orientés vers les relations humaines se montrent plus efficaces que les leaders orientés vers la tâche dans les situations modérément favorables ou modérément défavorables. Un leader sera considéré comme « apparié » (*in match*) lorsqu'il se trouve dans une situation où le modèle prédit une bonne performance du groupe et « non apparié » (*out of match*) dans le cas où il est dans un groupe avec une mauvaise performance (Fiedler et Chemers, 1984).

B. Le modèle en Trois Dimensions de Reddin (1986)

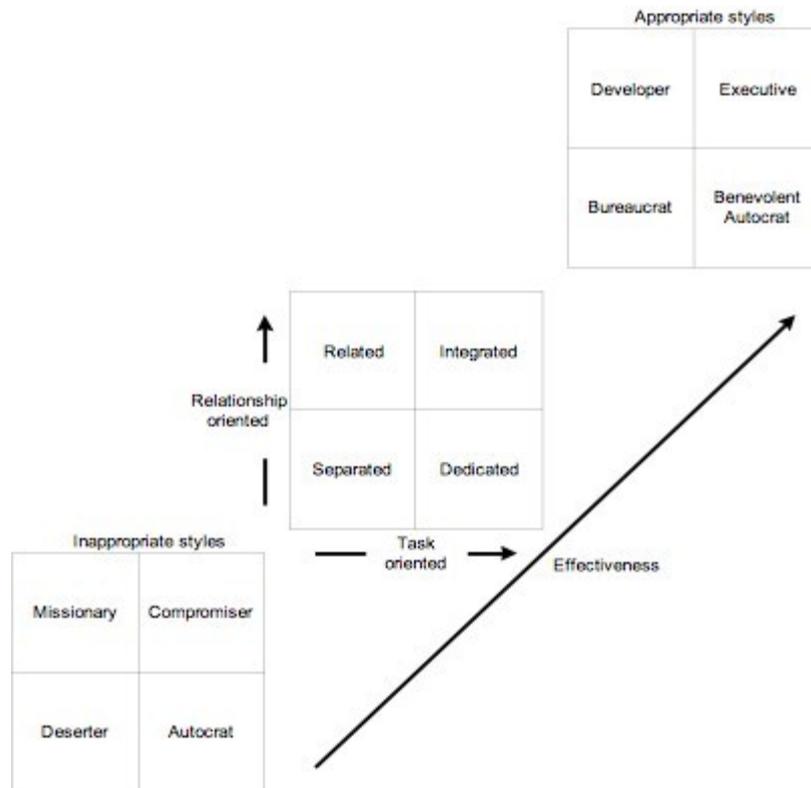
Reddin propose un modèle (Graphique T.3) basé sur la grille managériale de Blake et Mouton (1978).

Son modèle vise avant tout à proposer aux managers un outil simple et pratique centré sur l'efficacité. Les quatre styles de base sont des styles de management dont aucun ne serait efficace dans l'absolu mais deviendrait efficace selon les situations (Reddin, 1987). Aussi, Reddin ajoute un facteur supplémentaire aux dimensions Orientation vers la tâche et Orientation vers la personne : l'efficacité.

De la sorte, chacun des styles de base présente un versant efficace et un versant non-efficace.

Les quatre styles de base sont :

- ♣ Le séparé (*Separated*) présentant une faible orientation vers la tâche et vers les relations,
- ♣ Le Dévoué (*Dedicated*) ayant une forte orientation vers la tâche,
- ♣ Le Connecté (*Related*) ayant une forte orientation vers le relationnel et
- ♣ L'Intégré (*Integrated*) ayant une forte orientation et vers la tâche et vers le relationnel.



Graphique T.3 Modèle en 3D de Reddin

Reddin W. J. (website) English REFERENCE Reddin's 3-D Leadership model. OMG (Organisation for Managerial Development).

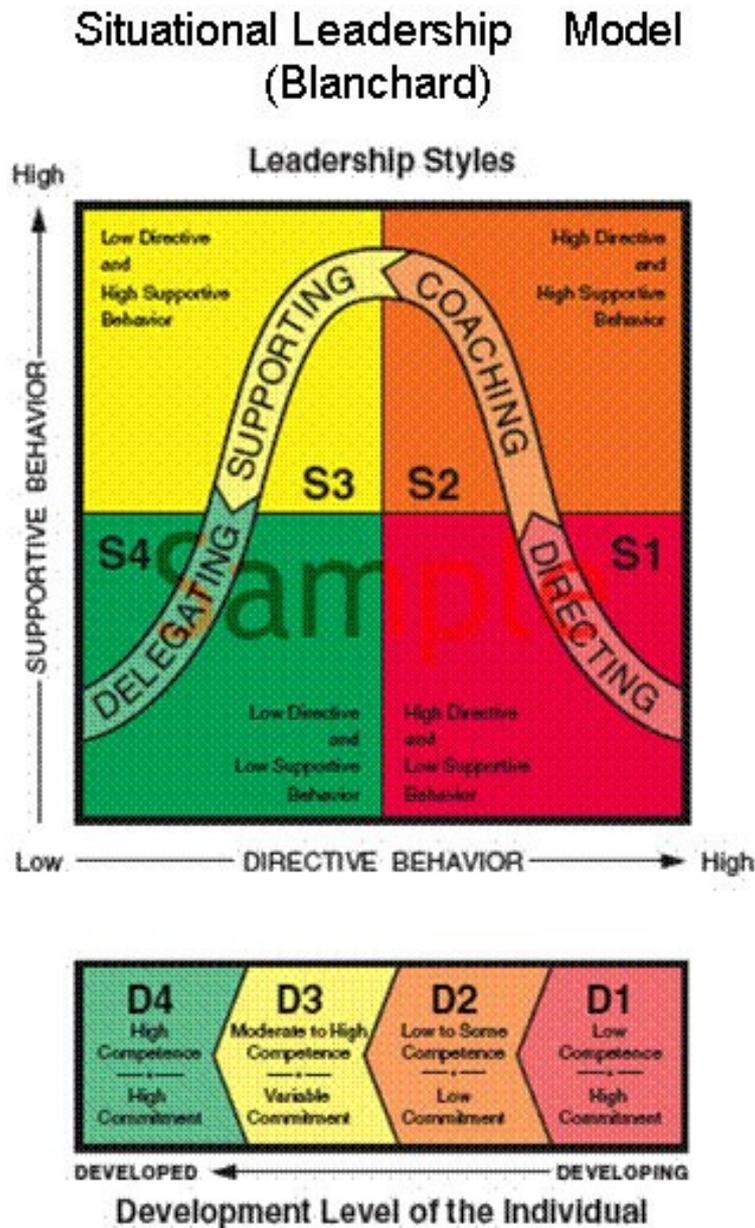
Sur le Graphique T.3, nous voyons au milieu les quatre styles de base, au-dessus et en-dessous le style approprié (permettant une bonne efficacité) et le style inapproprié (ne permettant pas d'être efficace) leur correspondant.

Reddin explique que l'efficacité du leader est une « fonction de l'accord entre un style de leadership et une situation donnée » (Reddin, 1986) Pourtant, le modèle ne permet pas à proprement parler de faire des prédictions quant à l'efficacité d'un style de leadership par rapport à un autre dans une situation donnée.

C. Le leadership situationnel, Hersey et Blanchard (1969)

La théorie de Hersey et Blanchard est en quelque sorte une extension du modèle en Trois D de Reddin (1986). En effet, les auteurs franchissent le pas que n'avait pas fait Reddin et associent plus directement la situation et le style de leadership.

Hersey et Blanchard (1969) intègrent dans leur modèle, en plus du style de leadership (orientation vers la tâche, vers les personnes), le degré de maturité des subalternes dans la situation considérée.



Graphique T. 4 Les Styles de leadership de Hersey et Blanchard.

Hersey, P., Blanchard K. H. et Johnson D. E. (2007) Management of organizational behavior: Leading human resources. 8th ed. Upper Saddle, NJ: Prentice Hall. p181

Ils définissent alors quatre styles de leadership qui s'associent aux quatre niveaux de développement des subalternes, évalués à partir de leurs compétences et de leur niveau d'engagement dans la tâche (Graphique T.4).

Le style (S1) **Orienter / Diriger** (*Directing*) se caractérise par un fort engagement sur les tâches et un faible engagement pour les relations. Les leaders définissent les tâches de chacun et les supervisent étroitement.

Le style (S2) **Vendre / Entraîner** (*Coaching*) se caractérise par un engagement fort aussi bien sur la tâche que la relation. En effet, comme les suiveurs présentent certaines compétences, un comportement centré uniquement sur la tâche risquerait d'anéantir leur motivation. Il est donc nécessaire de se centrer aussi bien sur la tâche (afin de leur permettre d'acquérir des compétences complémentaires) que sur la relation.

Le style (S3) **Participer / Soutenir** (*Supporting*) est faiblement centré sur la tâche et fortement sur la relation. Confronté à des suiveurs compétents et capables de réaliser le travail, le leader ne doit plus se focaliser sur la tâche mais s'efforcer d'augmenter la motivation et le bien être des suiveurs afin d'améliorer leur travail.

Le style (S4) **Déléguer / Observer** (*Delegating*) est faiblement centré sur la tâche et sur la relation. En effet, les suiveurs correspondant à ce style de leadership présentent une compétence et un engagement élevé, aussi le leader garde un œil plutôt distant sur leur travail.

D. La théorie « trajectoire-but » (*Path-goal Theory*)

Développée dans les années 1970 par Robert House (1971), cette théorie se propose de décrire comment les leaders motivent leurs subordonnés et les invitent à réaliser les objectifs visés.

Pour l'auteur, le leader y parviendra en utilisant les récompenses, en clarifiant les trajectoires et en enlevant les obstacles qui pourraient être rencontrés sur ces

trajectoires.

Selon la théorie, la réussite d'un leader dépendra de trois types de facteurs : le style de leadership, la personnalité des subordonnés et les caractéristiques de l'environnement (Tableau T.1).

Les quatre styles de leadership :

Le style « directif » (orienté vers la tâche) se caractérise par des règles et des procédures strictes fixées par le leader avec des critères de performance explicites.

Dans le style « centré sur les résultats » le leader met en place des objectifs stimulants et recherche l'amélioration de la performance en favorisant l'excellence.

Le style « supporteur » (orienté vers les relations) se caractérise par un climat convivial. Le leader est attentif au bien-être et aux divers besoins de ses subordonnés. Il les traite comme des égaux et avec respect.

Enfin, le leader au style « participatif » se caractérise surtout par la participation des subordonnés aux processus de prise de décision.

La personnalité des subordonnés :

Deux facteurs permettent à House de caractériser les subordonnés : leur locus de contrôle et leur aptitude perçue (*Self-perceived ability*). Ces deux facteurs présentent chacun deux modalités (interne / externe et élevé / faible respectivement).

Le style de leadership devra être adapté à ces caractéristiques des subordonnés tout en tenant compte des particularités des tâches à accomplir.

Les caractéristiques des tâches

Trois facteurs permettent de décrire les types de tâches :

- ⤴ Le niveau de structuration, d'organisation des tâches (ambigües, structurées)
- ⤴ Le système hiérarchique en place (système d'autorité formel ou plutôt informel)
- ⤴ L'environnement social du groupe (support social plus ou moins présent pour les collaborateurs dans le groupe)

Comportements du leader	Caractéristiques des subordonnés	Caractéristiques des tâches
Directif	Besoin de sécurité	<ul style="list-style-type: none"> - Ambiguës - Règles imprécises - Complexes
De support	Besoin social et d'affiliation	<ul style="list-style-type: none"> - Répétitives - Mécaniques - Non stimulantes - Banales
Participatif	Besoin d'estime de soi et de contrôler	<ul style="list-style-type: none"> - Ambiguës - Imprécises - Non structurées
Centré sur les résultats	Besoin de réalisation de soi	<ul style="list-style-type: none"> - Ambiguës - Stimulantes - Complexes

Tableau T. 1 Les liens entre les variables de la théorie « trajectoire-objectifs »

Northouse, P. G. (2004) Leadership: theory and practice, 3ème édition. Thousand Oaks: Sage Publications, 343 p.

E. Le modèle Normatif de Vroom et Yetton (1973), Vroom et Jago (1988)

Ce modèle repose sur la prise de décision et plus spécifiquement sur la participation des subalternes aux prises de décisions (Tableau T.2).

Il se propose d'indiquer aux leaders dans quelle mesure les subordonnés doivent prendre part à la prise de décision, et ce, en fonction d'un certain nombre d'éléments qui doivent être analysés par le leader.

Le modèle est constitué de trois éléments centraux : les stratégies possibles de prise de décision, les questions diagnostiques et les règles décisionnelles.

Cinq styles de leadership sont définis dans le modèle en fonction du niveau de participation des subalternes.

Style	Styles de prise de décision
A I	Le leader prend la décision seul
A II	Le leader recherche des informations puis prend la décision seul.
C I	Le leader consulte en face à face les membres du groupe puis prend la décision seul.
C II	Le leader consulte le groupe dans son ensemble puis prend la décision seul.
G II	Le leader partage le problème avec le groupe et la décision est prise ensemble (le leader intervenant au même niveau que n'importe quel autre membre du groupe)

Tableau T. 2 Les styles de prise de décision de Vroom et Yetton

Vroom V. H. (2003) Educating Managers for decision making and leadership. *Management Decision*, 41, 10, p972.

Ces cinq méthodes sont placées sur un continuum allant d'un style totalement autoritaire (A I) à un style participatif (G II).

Comme dans tous les modèles situationnels, le fonctionnement efficace dépend de la situation.

Pour définir le type de situation, et permettre au leader de choisir le style de leadership adéquat, les auteurs proposent un ensemble de sept questions (Vroom & Yetton, 1973 ; et dix dans Vroom & Jago, 1988) que le leader doit se poser. Après avoir répondu à ces questions (par oui ou par non) un arbre décisionnel permet de déterminer le type de participation à mettre en place afin de prendre la décision.

1. L'importance de la décision (l'implication du leader sera fonction de l'importance de la décision)
2. Le niveau d'engagement attendu des subordonnés (plus la mise en application de la décision demande un niveau élevé d'engagement des subordonnés, plus ceux-ci doivent être impliqués dans la prise de décision)

3. L'expertise du leader (le niveau de participation souhaitable des subalternes est inversement proportionnel à l'expertise du leader)
4. La propension à l'engagement (plus les subordonnés tendent à s'engager facilement dans le même sens que le leader, moins leur participation aux décisions est importante)
5. Le soutien / l'engagement du groupe face aux objectifs (moins les subordonnés sont engagés par les objectifs de l'organisation, plus leur participation aux décisions devrait être faible)
6. L'expertise du groupe (plus le groupe est compétent, plus grande devrait être sa participation aux décisions)
7. La compétence d'équipe (plus l'équipe est compétente et les individus prêts à travailler ensemble, plus les décisions peuvent leur être déléguées).

La spécificité du modèle de Vroom et Jago (1988) réside dans l'application de ces règles. En effet, les auteurs proposent un ensemble de règles décisionnelles permettant de choisir le bon niveau de participation à appliquer.

3. Les nouvelles approches du Leadership

En ce qui concerne les approches plus actuelles du leadership ce sont les théories reposant sur le leadership transformationnel qui semblent se distinguer.

Basées sur la notion de charisme, cette approche a été souvent associée aux théories des traits.

A. Le leader charismatique

On peut considérer que Max Weber a été le premier à parler de leader « Charismatique » (Weber, 1925, 1968). Dans sa recherche sur les sources d'autorité dans la société, Weber a proposé une typologie en trois types de leaders idéaux : le leader charismatique, le leader traditionnel et le leader légal-rationnel. Ainsi, d'après le modèle de Weber, contrairement aux deux autres types, le leader charismatique ne tire pas sa légitimité de la loi ou des règles. Pour l'auteur ce type de leader est donc souvent initiateur de changement.

Trice et Beyer (1986) proposent de résumer la théorie de Weber en cinq éléments qui permettent de caractériser les leaders charismatiques. D'une part il s'agit d'une personne extraordinairement douée et qui apparaît dans un contexte de crise sociale ou de situation désespérée. Pour faire face, le leader propose un ensemble d'idées qui sont des solutions radicales à la crise. Par ailleurs, le leader trouve un ensemble de suiveurs qui sont attirés par sa personnalité exceptionnelle et qui pensent qu'il est directement rattaché au pouvoir, pouvoir qui est validé par ses succès répétés.

1. La théorie de House (1977) du leader charismatique.

La théorie de House (1977) est une approche interactionniste du leadership. Pour l'auteur, le « leadership charismatique repose sur la relation entre le leader et les suiveurs » et les « leaders charismatiques transforment les besoins, les valeurs, les préférences et les attentes de ces subordonnés » (House, Spangler, et Woycke, 1991, p. 364). Par ailleurs, comme toute théorie interactionniste, les traits, les comportements et la situation sont pris

en compte.

Pour House, le charisme n'est donc pas considéré comme issu d'un certain nombre de traits de personnalité définis – mais les traits propres au leader seraient nécessaires à la mise en place de la relation charismatique entre le leader et les suiveurs.

Au niveau des traits, le leader charismatique présenterait quatre caractéristiques (House, 1977) : une tendance à la domination, un fort besoin de pouvoir (House et al., 1991; McClelland & Burnham, 1976) une forte confiance en soi et une grande conviction dans ses propres valeurs. Ce sont ces qualités qui créeraient chez le leader l'envie et la capacité de persuader et d'influencer les autres.

Par ailleurs, selon l'auteur, un environnement particulier est nécessaire pour l'apparition de ces leaders. En ce qui concerne House, les leaders charismatiques sont le plus susceptibles d'apparaître en situation de crise ou de danger.

Yukl (2002) et Northouse (2004) ont fait une revue de la littérature concernant les différentes caractéristiques du leader charismatique : ses traits de personnalité, ses comportements, et ses effets sur les subalternes (Tableau T.3).

Caractéristiques personnelles	Comportements	Effets sur les subalternes
• Dominant	• Démontre de la compétence	• Confiance dans l'idéologie du leader
	• Définit des objectifs	• Croyance en la similitude entre le leader et eux
• Désir d'influencer	• Exprime de la confiance	• Acceptent sans questionner
		• Affectionnent le leader
• Confiant	• Suscite l'engagement	• Obéissance
	• Se sacrifie	• Identification au leader
• Fortes valeurs	• Communique de hautes attentes	• Implication émotionnelle
		• Se fixent les objectifs plus élevés
• Innovateur	• Prend des risques personnels pour la cause défendue	• Accroissement de la confiance
		• Adoption des valeurs et croyances du leader

Tableau T.3 Caractéristiques personnelles, comportements, et effets du leader charismatique sur les subalternes.

Northouse, P. G. (2004) *Leadership: theory and practice, 3ème édition*. Thousand Oaks: Sage Publications, p78.

Un ensemble d'études a pu montrer la réalité de cette théorie (House et al., 1991; Howell & Higgins, 1990). Ainsi, Podsakoff et ses collaborateurs (1990) ont montré que les managers présentant une vision, montant l'exemple et communiquant de fortes attentes étaient ceux-là mêmes qui suscitaient le plus de confiance, de loyauté et de motivation de la part de leurs employés.

Toujours en partant des caractéristiques du leader charismatique, Conger et Kanungo (1987) ont proposé une théorie de l'attribution du charisme.

2. La théorie de l'attribution du Charisme de Conger et Kanungo (1987)

Conger et Kanungo (1987, 1997) ont développé un modèle centré sur certaines des dimensions charismatiques au sein des organisations. D'après les auteurs, les leaders charismatiques seraient créés par un mécanisme d'attribution des subalternes, basé sur leur perception du comportement de leur leader. Ainsi, le comportement observé du leader serait interprété par les subalternes comme une expression de son charisme. Ce leader charismatique se distinguerait des autres types de leaders par sa capacité à formuler et à inspirer une vision ainsi que par des comportements et actions qui donneraient l'impression qu'ils sont eux, tout autant que leur mission, exceptionnels (Conger & Kanungo, 1994).

Un des points majeurs de la théorie est de présenter les comportements du leader selon leur potentialité à provoquer l'attribution de la part des subalternes de caractéristiques charismatiques. Ainsi, les subalternes choisiraient de suivre ce leader non seulement du fait de son statut formel, mais aussi de la perception « d'extraordinaire » qu'ils ressentiraient.

Conger et Kanungo (1997) décrivent trois étapes dans la mise en œuvre du leadership (étapes qui semblent linéaires, mais se produisent le plus souvent de manière concomitante ou mélangée). A chacune de ces étapes les caractéristiques du leader, et donc les qualités attribués au leader vont varier.

La première étape est celle de l'évaluation de l'environnement, lors de laquelle, le leader est perçu comme voulant sortir du *statut quo* et comme étant particulièrement sensibles aux contraintes de l'environnement et aux besoins des subalternes.

Dans la seconde étape – celle de la formulation de la vision – le leader est perçu comme quelqu'un capable de proposer une vision idéalisée pour l'avenir, tout en évoquant les étapes de sa réalisation.

Enfin, dans la troisième étape, celle de l'implémentation de la vision, le leader est vu comme quelqu'un d'exemplaire, capable de prendre des risques personnels pour mettre

en place sa vision des choses.

D'après les auteurs, un leader sera perçu d'autant plus charismatique qu'il présentera une vision (différente de la situation actuelle et ayant un objectif de changement) qu'il encadrera et gèrera tout le long du processus (Conger, 1991). Ces leaders présenteront une sensibilité marquée à leur environnement et aux besoins des membres, et feront preuve de prises de risques personnelles dans le but de réaliser leur vision. Enfin, plus ils feront preuve de comportements inhabituels dans la réalisation de leurs objectifs, plus ils inspireront des qualificatifs charismatiques chez leurs subordonnés (Conger, 1991).

De même que dans la théorie de House (1977), Conger et Kanungo (1994 ; Conger, 1991) insistent sur la spécificité de la situation. Ainsi, pour les auteurs, les situations de crise ou de désenchantement sont les plus à mêmes de faire apparaître ce type de leader – bien que contrairement à House, pour Conger la crise n'est pas obligatoire.

3. Théorie du Self-Concept de Shamir (1991; 1993)

Introduite dans les années 90, la théorie du *Self-Concept* prolonge le concept de House du leader Charismatique en décrivant plus spécifiquement le processus à l'œuvre dans la phase d'influence. Ainsi la théorie du *Self-Concept* est une théorie motivationnelle qui se propose d'expliquer le processus par lequel les comportements d'un leader charismatique entraînent de profonds effets transformationnels sur ses suiveurs.



Dans sa théorie Shamir et ses collaborateurs (Shamir et al., 1993) expliquent que le leader charismatique agit sur ses suiveurs en engageant profondément leur concept même d'eux-mêmes (*self-concept*) (Graphique T.5). Ainsi, la mission présentée par le leader devient une part intégrante du suiveur, dans sa définition de lui-même.

Graphique T.5 Théorie du Self-Concept de Shamir

Shamir B., House, R. J., Arthur M. B. (1993) The motivational effects of charismatic leadership: a self-concept based theory. *Organization Science*, 4 (4), p581.

Les leaders charismatiques sont donc capables, avec un certain nombre de comportements qui leur sont spécifiques, d'activer des motivations propres à la définition de soi des suiveurs, ce qui entraîne chez ceux-ci des comportements particuliers.

B. Le leadership transformationnel

Vers la fin des années 70, une nouvelle conceptualisation émerge des études sur les leaders charismatiques : le leadership transactionnel et transformationnel.

1. La théorie des échanges Leader – Suiveur

Approche développée par Graen (1972) dans les années 70, la théorie LMX (*Leader-Member Exchange Theory*) centre la compréhension du leadership sur la relation entre le leader et le suiveur (Schriesheim, Castro, & Cogliser, 1999). C'est donc dans l'étude de la qualité, de l'efficacité puis de la spécificité de chacune de ces relations individuelles et dyadiques que le leader et son style peuvent être définis.

Aussi, malgré un certain nombre de désaccords concernant les dimensions impliqués dans la description et l'appréhension de cette relation (Schriesheim et al., 1999), les auteurs sont majoritairement d'accord sur le fait qu'il s'agit d'une approche de l'échange. Approche s'intéressant surtout à la quantité des relations inter-personnelles (entre le leader et le suiveur) ainsi qu'à la loyauté existant dans cette relation (Dansereau, Graen, & Haga, 1974; Graen & Cashman, 1975).

Enfin, l'importance de l'attention accordée au soutien et à la sensibilité sont également pris en compte par la théorie LMX, l'objectif principal de cette approche étant d'isoler les zones de difficulté dans la relation dans le but de l'améliorer afin d'accroître la réussite et l'efficacité de l'équipe.

2. Le leadership Moral de Burns (1978)

La théorie de Burns repose d'une part sur la théorie du leadership et de l'autorité de Weber (1948) et sur les stages de développement moral de Kohlberg (1964).

La théorie de Kohlberg propose trois niveaux de développement moral.

- ⤴ Au premier niveau (**moralité pré-conventionnelle**), le comportement est surtout déterminé par l'obéissance et par des récompenses et sanctions externes.
- ⤴ Au second niveau (**moralité conventionnelle**) la motivation se situe dans la conformité aux attentes de la société.
- ⤴ Enfin au troisième niveau (**moralité post-conventionnelle**) le comportement de l'individu obéit à des valeurs et à des principes universels.

En définissant le leadership, Burns (1978) veut séparer « la pure et simple possession de pouvoir du leadership proprement dit »

Aussi, pour présenter les divers leaders, Burns sépare dans un premier temps les leaders moraux et amoraux. Pour l'auteur les leaders amoraux sont motivés par des valeurs qui par définition les rendent non-leaders mais possesseurs de pouvoir, ils ne peuvent donc être ni transactionnels ni transformationnels.

Pour définir les leaders moraux, Burns précise que ceux-ci doivent « permettre aux gens de devenir plus conscients de leur ressenti. Le leader se doit de ressentir leurs besoins si profondément, de définir leurs valeurs si clairement qu'ils puissent être traduits en actions déterminées. » (notre traduction, Burns, 1978, p. 43-44¹)

Les leaders moraux peuvent être transactionnels ou transformationnels. Le leadership transactionnel s'installe lorsqu'un leader utilise des valeurs formelles (conditionnelles) pour diriger les suiveurs (parmi celles-ci nous pouvons trouver l'honnêteté, la responsabilité, l'intégrité...). Le leadership transformationnel, quant à lui, s'installe lorsque les valeurs du leader sont transcendantes telles que la liberté, la justice, l'égalité.

La taxonomie proposée par Burns présente donc d'une part les leaders amoraux (motivés exclusivement par l'utilisation du pouvoir) et d'autre part les leaders moraux qui sont soit transactionnels ou transformationnels (l'action transformationnelle étant présentée comme meilleure et plus aboutie que l'action transactionnelle).

¹ « The leader's fundamental act is to induce people to be aware or conscious of what they feel -- to feel their true needs so strongly, to define their values so meaningfully, that they can be moved to purposeful action. » (pages 43-44)

Contrairement à Burns (1978), Bass (Bass, 1985 tiré de Bass, Waldman, Avolio, & Bebb, 1987) dans le cadre de la conception multifactorielle du leadership insiste sur la différence conceptuelle entre le leadership transactionnel et transformationnel mais considère que ces deux dimensions apparaissent simultanément dans le répertoire comportemental du leader.

3. Le leadership transformationnel de Bass (1985; 1987)

Bass (1985), contrairement à Burns ne positionne pas le leadership transactionnel dans le domaine du moral, pour lui – seul le leadership transformationnel à une place dans cette catégorie et il le définit de manière multidimensionnelle. Aussi, un leader transactionnel va chercher à obtenir l'accord et la collaboration des subordonnés pour accomplir les tâches requises en usant de moyens externes (renforcements / punitions). L'accomplissement de ces tâches est basé sur les récompenses (renforcement positif des comportement permettant d'atteindre les objectifs) et le management par exception (comportement du leader basés sur la coercition et la punition des subordonnés, vis-à-vis de leurs erreurs). Deux types de management par exception sont présentés par Bass : MBEA (*Management by Exception – Active*) le management actif par exception où l'erreur est recherchée activement et MBEP (*Management by Exception – passive*) le management passif par exception où les leaders attendent d'apprendre les erreurs.

Le leader transformationnel se différencie, quant à lui, fondamentalement de celui-ci, dans la mesure où il n'a pas recours aux renforcements matériels. Il possède par ailleurs un certain nombre de caractéristiques qui lui sont propres :

« L'influence idéalisée » (*Idealized Influence*) est ainsi définie comme le fait d'être un modèle charismatique pour les autres, Bass tend d'ailleurs à simplement qualifier cette influence par le charisme et place cette caractéristique parmi les plus importantes. La « motivation inspirée » (*Inspirational Motivation*) implique l'articulation d'une d'une vision claire, attractive et inspirée pour les suiveurs.

Bien que ces deux dimensions soient conceptuellement distinctes, la recherche a montrée des liens forts entre les deux, elles sont donc souvent associés dans la pratique (Bass, 1998). « La stimulation intellectuelle » implique de stimuler la créativité des suiveurs en remettant en question leurs certitudes et le *statut quo*. Enfin, la « considération individuelle » est similaire à la dimension Considération telle que définie dans les théories des écoles de l'Ohio et du Michigan, et implique de porter de l'intérêt et de soutenir les besoins individuels des suiveurs. Enfin, contrairement à sa définition d'origine, la dimension de « considération individuelle » de Bass (1995) se différencie en se focalisant plus sur le développement personnel des suiveurs et moins sur la prise de décision participative.

Enfin, pour compléter les types de leaders, Bass ajoute aux leaders Transactionnels et Transformationnels, un style Laissez-faire qui, comme pour ses prédécesseurs, correspond plus à une absence de leadership qu'à un style de leadership à proprement parler

Pour développer ce modèle, Bass (1990) a utilisé un outil d'évaluation du leadership, le *Multifactor Leadership Questionnaire* (MLQ) – qui sera présenté plus en détail dans le chapitre trois (Chapitre 3.B.1).

En étudiant les effets de ces différents styles de leadership sur la satisfaction et la performance Bass positionne le leadership Transformationnel comme le meilleur des styles possible, le plus avancé et le plus efficace.

Dans le sillage des travaux de Burns et de Bass, plusieurs modèles de leadership transformationnel ont été développés. Parmi ceux-ci, l'un des plus populaires est celui de Kouzes et Posner.

4. James Kouzes et Barry Posner (1987) : les meilleurs comportements de leadership

Sur la base d'enquêtes exhaustives auprès de leaders-managers, Kouzes et Posner (1987) ont identifié cinq facteurs (types de comportements) qui décrivent les leaders exceptionnels. Ces derniers seraient, selon les auteurs, à mêmes de :

1. Remettre en question les processus (chercher de nouvelles opportunités, expérimenter et prendre des risques calculés)
2. Inspirer une vision partagée
3. Encourager les autres à agir
4. Tracer la voie – en donnant l'exemple par ses propres comportements
5. Reconnaître les contributions de chacun

A la suite de ces résultats, Kouzes et Posner (1993) ont élaboré un outil d'évaluation du leadership : *Leadership Practices Inventory (LPI)*, qui sera abordé plus en détail dans le chapitre trois (Chapitre 3.B.3).

- 1.

5. Comparaison entre les modèles

Sashkin (2004), auteur du modèle du leader Visionnaire, a fait une méta-analyse sur les divers modèles de leadership transformationnel. Il a conclu que les trois traits suivants sont ceux qui reviennent le plus souvent et sont jugés les plus importants dans tous les modèles étudiés :

1. La Capacité à articuler une vision
2. La confiance en soi
3. La capacité à favoriser *l'empowerment* (responsabilisation).

De même, il a isolé trois dimensions comportementales des leaders qui sont, elles aussi, partagées par la majorité des approches :

1. La capacité du leader à communiquer une vision
2. L'habilité du leader à créer des opportunités *d'empowerment* (responsabilisation)
3. La capacité du leader à montrer de la considération et du respect à l'endroit de ses

subalternes.

Pourtant, un certain nombre de problèmes méthodologiques ont immiscé, petit à petit, des doutes sur la validité de construit du concept de leadership transformationnel. Ainsi, d'autres auteurs (Kharti, 2005) proposent des structures différentes de celles initiées par Bass, dont certaines reviennent vers la théorie du leadership charismatique.

Ainsi, Kharti (2005) proposent que les quatre composantes postulées par Bass soient remplacées par deux plus centrales du nouveau construit théorique : le Charisme et la Vision. Ainsi donc, selon cette nouvelle approche, les leaders se trouvent catégorisés dans une des quatre catégories exclusives, basées sur leurs niveaux de charisme et leurs habilités à être visionnaires.

6. La théorie Intégratrice du leadership de Chemers (1997)

Dans sa conception unificatrice, Chemers propose d'intégrer les divers courants théoriques. Ainsi, les approches par trait, situationnelles, de contingences, transformationnelles et humanistes se trouvent conjuguées.

Ainsi, bien que certaines de ces approches postulent l'existence d'un meilleur type ou style, alors que d'autres considèrent qu'il n'y a pas de meilleur style dans l'absolu et que plusieurs bon styles existent (Skipton, 2003) Chemers propose un cheminement d'ensemble englobant ces différences.

Ainsi, dans sa théorie Intégratrice, l'auteur expose un chemin causal qui débiterait avec des caractéristiques personnelles du leader et les demandes de la situation (*zone of self-deployment*), puis des questions inter-personnelles apparaîtraient créant ainsi une zone centrée sur les relations transactionnelles (*the zone of transactional relationship*). Enfin, la mise en action des suiveurs ferait entrer les rapports dans la dernière zone: celle du déploiement de l'équipe (*the zone of team deployment*).

Aussi, pour réunir les théories, Chemers applique un principe d'appariement (*matching*) entre les efforts de leadership, les caractéristiques du leader et de les demandes de la

zone.

La zone du « **déploiement de soi** », demande principalement au leader de se centrer sur la gestion de l'image (Skipton, 2003). Dans cette zone, il s'efforce de faire correspondre les attentes de ses subordonnés et leurs perceptions du leader. A ce niveau, l'intérêt principal est donc de repérer (en se basant sur les perceptions des subordonnés) ce qui caractérise, d'après eux, le plus le leader (en termes de traits de personnalité et de situation).

La zone de « **relations transactionnelles** » est quant à elle centrée sur la dynamique et les inter-relations entre leader et subordonnés. La fonction principale du leader est donc de développer cette relation tout en maintenant et développant la motivation et l'engagement des subordonnés. D'après Chemers, cette zone a été le principal centre d'étude des courants transformationnels et de contingence.

Enfin, la zone du « **déploiement de l'équipe** » correspond à une phase d'utilisation et d'optimisation des ressources. La réussite du leader se traduira donc par la quantité de travail, d'efforts supplémentaire, de productivité des équipiers – variables d'effet que nous aborderons plus en détail dans le chapitre suivant.

5. L'approche clinique du leadership

L'approche clinique du leadership se place dans la suite logique des conceptions transactionnelles, transformationnelles et charismatiques du leadership. Elle vise à proposer un niveau complémentaire de lecture que ces approches n'ont pas jugé bon d'ajouter. Cette démarche clinique vise donc à approfondir la compréhension du rôle complexe et interactif du leader dans la société (Kets de Vries, 1993).

L'orientation clinique du leadership trouve son origine dans des théories aussi diverses que la psychanalyse, la théorie cognitive, la psychologie du développement ou l'approche systémique. La réunion de toutes ces approches ne vise qu'un seul objectif : comprendre la relation dynamique à l'œuvre entre le leader et ses suiveurs. (Kets de Vries et al., 2004).

Ainsi, l'objectif des tenants de cette approche sera de déconstruire le « théâtre intérieur » du leader (Kets de Vries, Korotov, et Florent-Treacy, 2007) afin de mettre en lumière les éléments fondamentaux à l'œuvre dans les comportements, pour le cas échéant, l'aider à effectuer des changements en profondeur et résoudre ces fonctionnements pathologiques spécifiques aux leaders dysfonctionnels (Kets de Vries, 2003a).

Aussi, contrairement aux autres approches, celle-ci ne se limite pas aux aspects cognitifs et comportementaux, mais prend également en considération les aspects émotionnels.

Dans cette approche le leader efficace joue deux types de rôles : un rôle charismatique et un rôle architectural.

Une part importante du rôle charismatique est en lien avec l'habilité du leader à donner une vision dynamique de l'avenir de l'entreprise – dimension également partie prenante des approches transactionnelles. Le leader se doit de donner une direction à prendre, de parvenir à mobiliser l'ensemble des salariés autour de l'objectif mais aussi de susciter la confiance quant à ses compétences de leader. Par cette vision, le leader parvient à donner un sens et à forger des liens entre lui-même et les autres, de même qu'à bâtir une identité fédératrice pour le groupe.

La capacité de déléguer le pouvoir est également un élément important du rôle de chef charismatique, tout autant que son habilité à motiver ses subordonnés.

Pour que cette vision puisse voir le jour et prendre corps le leader doit repenser les systèmes de récompenses et de contrôle afin d'encourager les types d'attitudes et les comportements qu'il souhaite voir se développer. Aussi, il importe que le leader réorganise le projet même de l'entreprise de sorte à ce que les valeurs de celui-ci s'harmonisent avec l'idéologie de fond de l'entreprise.

Ainsi, Manfred Kets de Vries (2003b) définit cinq grandes stratégies qui seraient génératrices de sens pour les salariés au sein d'une entreprise :

✦ Susciter le sens de la mission

En énonçant une vision claire de l'organisation à venir et en présentant des objectifs précis tout en incitant les salariés à s'engager dans cette voie et à y persévérer, les

managers renforcent la cohésion du groupe et donnent un sens aux objectifs collectifs.

✧ Susciter le sens de l'autodétermination

En renforçant la possibilité des employés d'effectuer des choix et de peser sur les décisions, les managers permettront aux employés de ne plus se sentir simplement spectateur mais également acteur de l'entreprise.

✧ Susciter un sentiment d'importance

En déléguant le pouvoir, les managers peuvent permettre à chaque employé de sentir le rôle spécifique qu'ils jouent dans le destin collectif de l'organisation. Cela permet également de renforcer leur sentiment d'utilité en donnant de l'importance à leurs contributions individuelles.

✧ Souligner le sens et la valeur de la compétence

Encourager l'apprentissage et le développement personnel permet aux employés de trouver un exutoire à leurs besoins d'exploration. Aussi, cela permet à chacun de prendre conscience de sa propre progression et de ses compétences.

✧ Développer le sens des valeurs partagées

Mettre en avant des valeurs telles que le travail d'équipe, la sincérité, l'orientation client, la responsabilisation ou l'ouverture au changement permet de centrer les valeurs de l'entreprise autour de grands idéaux partagés et c'est une des caractéristiques des entreprises qui réussissent.

Dans ce chapitre nous avons montré le passage d'une définition du leadership unitaires tenant compte d'un seul facteur : uniquement défini par la personnalité, par la situation ou seulement par le style du leader, à une définition plus interactionniste. Ces approches ont d'abord proposé des modèles tenant compte du style du leader et des caractéristiques de la situation, puis plus complexes, les caractéristiques des subordonnées sont venues compléter ces théories dans le but de proposer un cadre de référence permettant de prédire l'efficacité optimum d'un leader donné en fonction des circonstances au sens large.

C'est alors que l'ancienne notion de leader Charismatique, déjà présente dans les années 1940 a été remise au goût du jour, donnant naissance aux modèles centrés sur les relations leader-subordonné et plus spécifiquement sur l'effet des leaders sur leurs subordonnés. Ainsi, la grande famille des théories du leadership charismatique, transactionnel et surtout transformationnel sont apparues. Enfin, partant du constat du manque de profondeur de ces approches, certains auteurs ont proposé une conception plus clinique du leadership. Celle-ci se propose d'intégrer aussi bien des variables comportementales, cognitives que émotionnelles au sein de son modèle.

De fait le leadership apparaît donc à présent comme un construit complexe, varié et plus nuancé qu'il ne paraissait lors des premières définitions. Aussi, nous allons développer une compréhension plus différentielle du leadership en mettant en lumière les liens qu'il entretient avec les dimensions de la personnalité ou de l'intelligence émotionnelle. Dans tous les cas, toutes ces définitions visent un seul et même objectif : comprendre et décrire le leader efficace. Mais qu'en est-il au juste de l'efficacité des leaders ? Comment se conçoit-elle ? Comment pouvons nous la mesurer ?

En ce qui concerne la mesure, lors de notre présentation des différents modèles du leadership nous avons également évoqué les nombreux outils de mesure élaborés, souvent, en parallèle de la théorie et visant à fournir des éléments empiriques de validation de celles-ci. Comme nous le verrons dans les chapitres à venir, ces outils se sont peu à peu transformés en tests psychométriques à visée diagnostique pour certains et se proposent aujourd'hui d'évaluer les managers sur leurs qualités de leader.

Chapitre 2 : Le Leadership, une dimension pas si indépendante

Nous avons vu dans le chapitre précédent la diversité des approches et de définitions du leadership. Mais toutes ces approches ont un point commun : elles cherchent à définir le leader efficace. En effet, au-delà des traits ou des styles évoqués, les modèles recherchent avant tout à mettre en lumière ce qui caractérise un leader efficace. Aussi, dans la première partie de ce second chapitre, nous allons nous intéresser à la manière dont un leader peut-être considéré comme efficace. Nous allons donc chercher à comprendre comment ont été définis et évalués l'efficacité ou le succès d'un leader. Pour ce faire nous allons étudier un ensemble de variables d'effets utilisés pour évaluer l'efficacité du leader.

Puis, dans la seconde partie du chapitre, nous verrons comment l'approche différentielle s'intéresse aux liens que le leadership entretient avec d'autres variables psychologiques comme la personnalité ou l'intelligence émotionnelle. Car comme nous l'avons vu, ce concept est loin d'être unitaire et englobe dans ses théorisations récentes un grand nombre de variables allant des dimensions cognitives aux facteurs émotionnels.

1. Les variables d'effets et l'efficacité du leadership

Du fait de la diversité des éléments que recouvre le leadership et de sa nature profondément « relationnelle », beaucoup de chercheurs ont défini les effets du leadership à travers un ensemble de variables très larges.

Ces variables d'effet vont de la perception subjective de l'efficacité des leaders ou du groupe (Carless, Wearing, & Mann, 2000; Garman, Tyler, & Darnal, 2004; Judge & Bono, 2000; Vandenberghe, Stordeur, & D'hoore, 2002) à des critères économiques ou de performance (Avery, Tonidandel, Griffith, & Quinones, 2003; Carless et al., 2000; Connelly et al., 2000; Geyery & Geyery, 1998; Khuntia & Suar, 2004; Marshall-Mies et al., 2000; Misumi, 1995) Nous allons voir dans la suite de cette partie comment les chercheurs ont opérationnalisé l'efficacité du leadership et surtout comment ils l'ont mesurée.

A. Perception de l'efficacité des leaders

Afin d'estimer la perception de l'efficacité des leaders, certaines études ont fait appel à des juges (par exemple un groupe de superviseurs) pour obtenir un classement des managers selon leur efficacité (Garman et al., 2004). D'autres ont utilisé des questionnaires spécifiques pour appréhender cette variable.

Ainsi, Ragins (1989) a choisi de développer une échelle pour mesurer la perception par les subordonnés de l'efficacité de leur manager. D'après l'auteur, il est difficile de mesurer l'efficacité perçue des leaders à partir de certains comportements spécifiques car une part importante de l'efficacité de ceux-ci résulte de leur capacité d'influence.

De ce fait, l'auteur a fait le choix d'utiliser des items « larges » mesurant l'efficacité. Il a construit une échelle composée de cinq items tels que : « Mon leader dispose de grandes capacités de leadership » à laquelle les subordonnés doivent répondre sur une échelle de type Likert en sept points allant de « pas du tout d'accord » à « tout à fait d'accord ».

Ces items ont montré des corrélations significatives et positives avec des items plus spécifiques représentant 46 comportements efficaces du leader (.65). Cependant, aucune donnée n'est disponible quant à la consistance de ces items. Il est également à noter que cette échelle a été construite spécifiquement pour une recherche donnée (Ragins, 1989).

Carless en 2000 a enrichi l'échelle de Ragins. Dans leur étude, Carless et ses collaborateurs (2000), ont proposé une dimension complémentaire pour évaluer la perception de la qualité du travail, de la productivité et de l'adaptabilité du manager.

Les items de cet outil sont basés sur le travail conceptuel de Mott (1972) sur les caractéristiques des organisations efficaces, l'échelle est composée d'items tels que : « Il/Elle produit un travail de grande qualité » avec des réponses sur une échelle de type Likert en cinq points allant de « rarement ou jamais » à « très souvent ou toujours ». Cette nouvelle échelle n'a pas été validée au sens psychométrique.

Enfin, dans d'autres études, la formulation de certains items issus de questionnaires de leadership a été transformée pour mesurer la perception que les subordonnés ont de l'efficacité de leur leader. Il va sans dire, que ces échelles modifiées n'ont pas été validées en amont de leur utilisation.

Cette démarche utilisant les évaluations des subordonnés pour évaluer les qualités du leader sera abordée plus en détail dans le chapitre 4 à propos des évaluations par l'approche à 360°.

B. Efficacité et perception de l'efficacité du groupe

Au-delà de l'efficacité du leader, certains chercheurs ont tenté d'évaluer la qualité d'un style de leadership en s'intéressant à l'efficacité du groupe ou à la perception qu'en avaient les employés.

Deux catégories de variables ont ainsi été évaluées dans ce cadre : l'atteinte des objectifs et l'attraction de l'unité / de l'équipe.

Bass et Avolio, ont dès 1990, dans la forme Cinq du MLQ (*Multifactor Leadership Questionnaire*) ajouté deux dimensions propres aux effets du leadership : l'efficacité et la satisfaction. Le questionnaire de Bass et Avolio sera détaillé dans le chapitre 3 (Chapitre 3.3.B.1).

L'échelle portant sur l'efficacité est composée de quatre items mesurant l'efficacité perçue du leader, à travers des questions telles que : « A quel point votre supérieur est efficace lorsqu'il doit répondre aux besoins professionnels de ses employés ». Cette échelle a été utilisée à de nombreuses reprises dans des études aussi différentes que celles portant sur l'efficacité du leadership dans des entreprises du fret aérien (Hater & Bass, 1988) que sur les qualités de leadership des officiers de la marine (Waldman, Bass, & Yammarino, 1990)

Mais certains milieux ont leur propres spécificités, aussi, en fonction des milieux dans lesquels les études se sont déroulées différentes approches ont pu être mises en place pour estimer l'efficacité.

Ainsi, dans le cadre de leurs études sur le milieu médical, et plus spécifiquement sur les unités de soins intensifs, Shortell, Rousseau, Gillies, Devers et Simons en 1991 ont développé un questionnaire évaluant, d'une part, la capacité de l'unité à attirer et à garder des infirmières et docteurs de haut niveau et, d'autre part, le degré auquel l'équipe parvenait à achever sa mission de soin auprès des patients.

Pour mesurer l'efficacité de l'équipe, ces auteurs donc ont développé un outil complet en prenant des items issus de l'échelle de performance au travail d'Abrahamson (1985) qui portent sur : la prise de décisions, le respect des délais/échéances, la production d'un

travail de qualité suffisante, la production d'un travail en quantité suffisante, une bonne organisation du travail, la capacité de faire face aux situations conflictuelles, la confiance face au travail à faire.

Ils ont également demandé aux employés d'évaluer la qualité de la performance, la productivité, le coût et de la sécurité perçue dans leur travail (à travers des questions objectives). Enfin en utilisant des items issus du travail de Dansereau et collaborateurs (1984), ils ont évalué la fréquence avec laquelle l'employé fournissait une performance correspondant à la ligne directrice de son manager.

Cette échelle, parfois légèrement modifiée pour s'adapter à d'autres unités de soins, à été régulièrement reprises dans d'autres études (Vandenberghe, Stordeur, & D'hoore, 2002) pour estimer l'efficacité d'un service et le mettre en lien avec des styles de leadership différents.

Aussi, bien que les qualités métriques de l'échelle originale de Shortell et collaborateurs (1991) soient acceptables (Alphas de Cronbach supérieurs à .6 pour toutes les dimensions) ce questionnaire à très rarement été utilisée dans son ensemble. Beaucoup de recherches se sont limités à prendre quelques items de l'échelle. Qui plus est, étant spécifiquement conçue pour s'adapter aux unités de soins intensifs, elle a souvent été adaptée pour pouvoir être utilisée dans d'autres services hospitaliers. Or ces versions adaptées n'ont pas fait l'objet d'une validation.

C. Satisfaction des employés avec le leader / le travail

Tel que défini par Locke en (1976), « la satisfaction avec son travail est issue d'une réaction émotionnelle liée à la croyance que son travail doit remplir certains besoins et répondre à certaines valeurs. »

Aussi, comme les employés peuvent être satisfaits ou insatisfaits d'un grand nombre de paramètres liés à leur travail, cette question de la satisfaction a été abordée de manières différentes selon les auteurs.

Sparks et Schenk (2001) ont ainsi adapté quatre items de la sous-échelle générale de satisfaction au travail de Haclonan et Oldham (*Job Diagnostic Survey*; cité par Sparks et Schenk, 2001).

Vandenberghe et collaborateurs, dans leur grande étude de (2002)² ont abordé la satisfaction selon plusieurs angles. D'une part la satisfaction avec le leader a été mesurée à l'aide de deux items. Ceux-ci étaient ancrés à un bout de l'échelle avec le terme « très insatisfait » et à l'autre bout de l'échelle avec le terme « très satisfait ». D'autre part, la satisfaction générale avec le travail a été mesurée avec un item de Shortell et collaborateurs (1991): « Dans l'ensemble, à quel point êtes vous satisfait(e) avec votre travail? ». Les subordonnés devaient répondre sur une échelle d'accord de type Lickert en cinq points allant de « très satisfait » à « pas du tout satisfait ».

Comme nous l'avons mentionné précédemment (partie 1.B de ce chapitre), Bass et Avolio, ont proposé dès 1990 dans la forme cinq du MLQ (*Multifactor Leadership Questionnaire*) deux échelles destinées à mesurer le leadership à travers ses effets. Nous avons présenté l'échelle d'efficacité dans la partie 1.B de ce chapitre. La dimension satisfaction du MLQ vise à mettre au jour la satisfaction des employés avec leur manager. Elle est composée de deux items : « En général, à quel point êtes vous satisfait(e) de votre supérieur? » et « En général, à quel point trouvez vous que les méthodes utilisées par votre supérieur sont celles qui permettent le mieux au groupe d'atteindre le résultat escompté? ».

Malgré les nombreuses études de validation du MLQ, aucune information psychométrique n'est donnée concernant ces deux échelles (composées de deux questions chacune). Elles sont par ailleurs décrites comme des échelles complémentaires donnant des informations à visée descriptive.

D. Efforts supplémentaires des employés / implication dans le travail

Pour définir les efforts supplémentaires que les employés sont prêts à réaliser, Carless (2000) parle de la motivation que les subordonnés attribuent à leur leader.

Ainsi, dès 1985 Bass propose une échelle spécifique dénommée « *Extra Effort Scale* » avec des items tels que « Mon manager me motive à me dépasser. ». Elle est utilisée dans de nombreuses études pour mesurer l'effet du leadership (Bycio, Hackett, et Allen, 1995) telle quelle ou parfois légèrement modifiée pour être « plus spécifiquement reliée à la performance au travail » (Carless et al., 2000).

Nous pouvons noter, que dans les études où l'échelle est modifiée, aucune information psychométrique sur ses qualités n'est précisée.

² Réalisée auprès d'un échantillon de plus de 1000 infirmières belges.

Enfin, d'autres chercheurs ont utilisé l'échelle d'implication dans le travail de Lodahl et Kejnar (1965) afin d'estimer l'effet du leadership de leur manager sur l'implication au travail des salariés. L'échelle de Lodahl et Kejnar présentait lors de son développement et des études de validation associées, des qualités métriques acceptables, bien que limitées. Ainsi, les auteurs eux-même qualifient la consistance de leur échelle de « adéquate mais non élevée ³» et expriment des doutes quant à sa structure factorielle (traduction libre, Lodahl et Kejnar, 1965, p. 32).

Enfin, nous pouvons noter que cette échelle date de 1965 et qu'elle est utilisée telle quelle par les chercheurs. Aussi, pour nous prononcer sur sa pertinence et ses qualités, il serait nécessaire d'avoir des données psychométriques plus récentes.

E. Intention de quitter le poste

L'intention de quitter le poste peut être perçue comme une sous-dimension de la satisfaction avec son travail. Cependant, cette dimension est souvent associée au leadership. Ainsi, Vandenberghe et ses collaborateurs (2002) ont introduit cette variable dans leur étude pour estimer le résultat et l'efficacité du leadership des managers.

Les auteurs ont évaluée l'intention des employés de quitter leur poste grâce à 3 items portant d'une part sur l'intention de quitter ce poste précis et, d'autre part, sur l'intention de quitter la profession en général.

Pourtant tous les auteurs ne sont pas unanimes quant à cette variable d'effet. Pour certains, l'intention de quitter le poste ne serait pas forcément un signe de l'inefficacité du leader (Bycio et al., 1995). En effet, certaines dimensions du leadership transformationnel, associées de manière appropriée pourraient engendrer cet effet « indésirable ». Si l'on reprend la définition du leadership transformationnel, il est envisageable que des récompenses adaptées associées avec un leadership inspirationnel et un environnement présentant une dose suffisante de challenge, puissent augmenter chez certains employés le désir de quitter leur poste, pour évoluer (Bycio et al., 1995).

³ Adequate, but not high reliability.

F. Engagement envers l'organisation

Dans leur article de référence datant de 1990, Allen et Meyer proposent un modèle de l'engagement envers l'organisation structuré autour de trois axes :

- Une composante affective de l'engagement envers l'organisation qui se réfère à l'attachement, l'identification et l'implication émotionnelle de l'employé avec son entreprise.
- Une composante de continuité qui est liée à l'engagement moral basé sur le coût que l'employé associe avec le fait de quitter l'entreprise.
- Enfin, une composante normative qui correspond au sentiment d'obligation que ressent l'employé de rester dans l'entreprise.

Pour de nombreux chercheurs, l'engagement envers l'organisation est un résultat important du leadership exercé dans l'entreprise ou l'organisation, aussi, l'échelle d'Allen et Meyer est régulièrement utilisée dans sa totalité ou en partie (Bycio et al., 1995; Vandenberghe et al., 2002).

Deux autres échelles sont souvent utilisées pour la mesure de l'engagement envers l'organisation. Celle de Mowday et collaborateurs (1979) et des adaptations de l'échelle de Howell et Avolio (1988) s'intéressent à l'engagement envers différentes parties du travail.

Nous avons donc vu dans cette première partie les nombreux effets qu'un leader efficace peut provoquer au sein des organisations en général et sur ses subordonnés en particulier. Nous avons également entrevu les difficultés liées à la mesure de ces effets, mais aussi à la place que les subordonnés et autres collaborateurs prennent dans cette évaluation. Aussi avant d'aborder la mesure du leadership à proprement parlé (Chapitre 3) nous allons nous pencher sur les liens – parfois étroits – que le leadership entretient avec d'autres construits psychologiques. Pour mieux comprendre ce concept multidimensionnel, il importe – dans une démarche différentielle – d'ancrer le leadership dans les théories de la personnalité et d'explorer ses liens avec les dimensions conatives de l'intelligence émotionnelle.

2. Lien avec d'autres dimensions psychologiques

Comme cela se lit entre les lignes des différentes théories du Leadership – ce concept est loin d'être unitaire et semble par ailleurs entretenir des liens étroits avec d'autres variables. En effet, le leadership englobe – surtout dans les théories récentes – un grand nombre de variables qui semblent se rapprocher de certaines caractéristiques basiques de la personnalité

A. Les traits de personnalité

Depuis le début des années 40 les psychologues se sont intéressés aux liens entre les aptitudes au leadership et à la personnalité. Mann (1959) et Stogdill (1974) ont effectué en leur temps de grandes revues synthétiques du leadership – que nous avons évoquées dans le premier chapitre (Chapitre 1.A.1.) – dans lesquelles ils ont cité de nombreuses recherches étudiant les liens entre traits de personnalité et le leadership.

Ainsi, dans leur synthèse, les auteurs rapportent un ensemble de caractéristiques de personnalité ayant des liens positifs avec le leadership. Parmi les principales, nous retrouvons l'éveil, l'originalité, l'intégrité personnelle et la confiance en soi. Les dimensions d'agressivité, d'indépendance, d'objectivité, d'ajustement et de tolérance au stress sont, elles aussi, souvent citées.

Plus récemment, un regain d'intérêt a pu être observé dans l'étude des liens entre la personnalité et le leadership. Ainsi, en (1986), Lord et collaborateurs ont trouvé que la dimension dominance ainsi que la dimension Masculinité-Féminité permettait de prédire l'émergence du leadership.

Les émotions, surtout les émotions positives ont, elle aussi, été beaucoup étudiée comme un antécédent possible du leadership transformationnel. Ainsi la forte présence d'émotions positives chez le leader a été associée avec une meilleure performance (Wong & Law, 2002), avec la réussite (Staw, Sutton, & Pelled, 1994), l'efficacité du leader (George, 1995), le maintien de bonnes relations au travail (Staw et al., 1994) et une évaluations plus positive par les subordonnés des qualités de leur leader (Newcombe & Ashkanasy,

2002).

1. Les cinq dimensions fondamentales de la personnalité

La plupart des recherches récentes sur les relations entre leadership et personnalité s'inscrivent dans le cadre du modèle des traits de personnalité dit des *big five*. Nous allons donc rappeler brièvement son histoire et son rationnel.

Les recherches sur la structure de la personnalité ne sont pas récentes, dans l'antiquité, les philosophes tentaient déjà de classer les individus selon leur tempéraments. Ainsi, Cattell, vers la fin des années 50, a entrepris un travail pharaonique avec son équipe visant à isoler à partir essentiellement du vocabulaire, les dimensions fondamentales de la personnalité.

Cattell (1957) et son équipe ont employé 3 méthode pour collecter autant d'information que possible sur les caractéristiques des personnes. Les Données-L (*Life Data*) correspondent aux données issus de biographies et journaux intimes. Les Données-T (*Test Data*) correspondent aux données recueillies expérimentalement, en mettant des volontaires dans diverses situations émotionnelles et en récoltant leur réponses. Les Données-Q (*Questionnaire Data*) correspondent aux résultats issus des questionnaires construits par Cattell et son équipe en recensant tous les adjectifs présents dans le dictionnaire, pouvant s'appliquer à la description d'une personne. Ils ont ensuite soumis collecté des réponses à ce questionnaires auprès de volontaires qui ont utilisé ces adjectifs pour se décrire eux-mêmes et pour décrire d'autres personnes.

Ces résultats issues de toutes ces collectes ont été analysés grâce à la méthode d'analyse MAVA (*Multiple Abstract Variance Analysis*) pour identifier les construits latents permettant de grouper ces adjectifs. 19 variables latentes ont ainsi été mises en lumière lors de la première analyse, finalement réduits à 16 dans les analyses suivantes.

Ce n'est que vers la seconde moitié du vingtième siècle qu'une structure plus fondamentale encore à été mise à jour. Ce sont les auteurs Tupes et Christal (1961) qui sont le plus souvent considérés comme les premiers auteurs de cette théorie, et ce bien que leurs travaux aient porté sur une re-analyse des résultats de Cattell. Suivis par Norman (1963) qui lui aussi à ré-utilisé les matrices de Cattell. Ces auteurs ont proposé une première dénomination des cinq dimensions fondamentales de la personnalité :

- Extraversion / Surgence (*Extroversion / Surgency*)
- Agréabilité (*Agreeableness*)
- Conscience (*Conscientiousness*)
- Stabilité Émotionnelle (*Emotional Stability*)
- Culture (*Culture*)

Mais ce sont les travaux de Costa et McCrae (McCrae & Costa, 1985; 1987) dans les années 80 qui, travaillant eux aussi sur un modèle de personnalité fondamental, ont donné naissance au premier outil permettant de mesurer ces cinq traits de personnalité : le NEO-PI (*Neo Personality Inventory* – Costa & McCrae, 1985). Comparativement à Norman, Costa et McCrae ont ré-interprété le facteur Culture par la dimension Ouverture à l'expérience (*Openness to experience*) (McCrae & Costa, 1985).

Dès les premières études de validation du Neo-PI, ces cinq dimensions fondamentales ont été retrouvées lors d'analyses d'adjectifs (Goldberg, 1983), et d'autres questionnaires de personnalité comme le Block (1961) ou le California Q-set (McCrae, Costa, & Busch, 1986).

Depuis, de très nombreuses études ont montré la robustesse de cette structure à travers les cultures et les mesures. Les *Big Five* ont donc connu une popularité croissante au cours des années notamment parce qu'elles représentent des dimensions de personnalité vastes, qui se manifestent dans des traits plus spécifiques et que des outils permettaient de les mesurer aisément (McCrae & Costa, 1987; McCrae & John, 1992). Les auteurs définissent les cinq dimension comme suit :

- ♣ La première dimension – **Extraversion** – représente la tendance à être sociable, assertif, dynamique et à la recherche de stimulations.
- ♣ La seconde dimension – **Agréabilité** – correspond à une tendance à être gentil, aimable , confiant et chaleureux.
- ♣ La troisième dimension – **Conscience** – est composée de deux facettes : l'atteinte des objectifs et la confiance que l'on est prêt à nous accorder. C'est ainsi le trait des *Big Five* qui est le plus en lien avec la réussite professionnelle (Barrik & Mount, 1991).
- ♣ La quatrième dimension – **Stabilité Émotionnelle** – est le plus souvent définie par son pôle opposé, Neuroticism. Elle correspond à la tendance à l'anxiété, à la peur, à la dépression et c'est la dimension la plus associée à la satisfaction générale avec sa vie (Costa & McCrae, 1992).

- △ Enfin le cinquième facteur – **Ouverture** – représente la tendance à être créatif, imaginatif et ouvert à l'expérience.

Comme nous le verrons, c'est le Neo-PI qui est le plus souvent utilisé dans les études sur les liens entre leadership et personnalité.

2. Liens Big Five - Leadership

Dans les années 2000 Judge et Bono (2000), ont réalisé une grande étude de mise en relation des dimensions du leadership Transformationnel et Transactionnel avec les Dimensions des *Big Five*. Les participants de leur étude étaient engagés dans des programmes de leadership communautaires, mis en place et financés par ces Chambres de Commerce des régions du *Midwest* américain. Les participants ont reçu un ensemble de questionnaires à compléter par eux-mêmes, par leur supérieur et subordonnés.

Les dimensions de personnalité étaient mesurées à l'aide du *Neo-PI Révisé* (Costa & McCrae, 1992) tandis que les comportements de leadership transactionnel et transformationnel ont été évalués grâce au *MLQ-5x (Multifactorial Leadership Questionnaire - version cinq, Avolio, Bass, & Jung, 1995)*. De même, des questionnaires de satisfaction (satisfaction de leader avec son travail, satisfaction des subordonnés avec le leader et satisfaction du supérieur avec le leader) ont été administrés.

Dans le cadre de cette étude, les auteurs ont trouvé un lien direct et significatif entre la personnalité et le leadership. Plus particulièrement, la dimension Agréabilité serait pour eux un fort et consistant prédicteur de comportements de leadership transformationnel.

De même, Atwater et Yammarino (1993) en effectuant une étude dans le cadre militaire, se sont intéressés aux prédicteurs liés avec la perception du leadership militaire. Il ont pour ce faire, mis en relation des variables d'attributs personnels (telles que les traits de personnalité, les styles cognitifs...) avec l'évaluation des dimensions de leadership transactionnel et transformationnel qu'avaient les supérieurs et les subordonnés d'un leaders militaire. Les auteurs ont ainsi trouvé que le fait d'être chaleureux était un prédicteur significatif des comportements de leadership transformationnel.

Dans le même sens, Cavazotte et ses collaborateurs (2012) en s'intéressant aux effets de l'intelligence, de la personnalité et de l'intelligence émotionnelle sur le leadership transformationnel et la performance managériale ont eux aussi montré que le Neurotiscisme (pôle opposé de la stabilité émotionnelle) avait un effet négatif sur l'efficacité du leader. Pour ce faire ils ont interrogé des managers d'une grande entreprise énergétique brésilienne. Pour évaluer la personnalité les auteurs ont utilisé une échelle de Goldberg (1999) permettant de mesurer, entre-autre, les cinq traits des *Big Five*. Pour le leadership, ils ont utilisé la version portugaise du MLQ (Antonakis, Avolio, & Sivasubramaniam, 2003).

D'ailleurs pour Bass la stabilité émotionnelle serait prédictive de formes de leadership moins complexes que le leadership Transactionnel ou Transformationnel, comme le leadership Laissez-faire (Bass, 1998).

Enfin, Keller (1999) dans son étude portant sur le leadership implicite, s'est intéressé à l'impact des traits de personnalité, de l'estime de soi et de la conscience de soi des leaders sur le leadership implicite. Cette étude a porté sur des étudiants suivant un cours d'introduction au leadership. L'auteur a pu montrer que la dimension Agréabilité, était significativement reliée aux évaluations de la sensibilité des leaders.

Un certain nombre d'études se sont intéressés aux les liens entre personnalité et leadership dans le cadre militaire et ont mis en lien les traits de personnalité et le leadership (Lim & Ployhart, 2003; Ployhart, Lim, & Chan, 2001; Thomas, Dickson, & Bliese, 2001).

Lim et ses collaborateurs, ont étudié, entre autres, les liens entre l'évaluation par les subordonnés du leadership de leur chef et les traits de personnalité de celui-ci, tels que définis par le modèle des *Big Five*. Travaillant sur des groupes de combat des forces armées de Singapour, ils ont montré qu'en régressant les dimensions de leadership transformationnel sur les cinq dimensions de personnalité du modèle des *Big Five*, ils était possible d'expliquer 28% de la variance des évaluations du leadership transformationnel. Travaillant toujours avec les militaires, d'autres auteurs ont trouvé des liens forts entre la dimension Extraversion et des comportements de leadership transformationnel (Ployhart et al., 2001) et que cette dimension d'extraversion était un prédicteur significatif de la performance de leaders des cadets évalués dans le cadre de leur entraînement (Thomas et al., 2001).

Pourtant, Judge et Bono (2000) – en s'intéressant aux étudiants du *Midwest* - ont trouvé l'extraversion seulement marginalement prédictive, ces mêmes auteurs ont d'ailleurs fortement insisté sur le fait que dans de nombreuses recherches antérieures l'extraversion présentait souvent des résultats mitigés avec le leadership.

La dimension Conscience a été, elle aussi considérée comme une dimension prédictive de l'émergence du leadership (Judge & Bono, 2000) et dans certains cas, cette dimension a même été liée à l'efficacité du leader (Cavazotte et al., 2012).

3. Le modèle HEXACO de la personnalité

Depuis son début, le modèle de la personnalité en cinq facteurs a souvent été contesté, ainsi Goldberg (1993; 1981), qui comme Cattell, a travaillé sur une approche lexicale de la personnalité, s'est penché sur les mots (plus spécifiquement les adjectifs) issus de la langue anglaise. Cette démarche lui a permis de mettre en lumière cinq grands traits de personnalité, il a donc de ce fait rejoint le grand groupe de recherche sur le modèle de la personnalité en cinq traits. C'est en prolongeant ses études linguistiques à d'autres langues que Goldberg et son équipe ont trouvé des différences (Ashton et al., 2004; Ashton, Lee, & Son, 2000). Basé sur ces résultats les chercheurs ont proposé un nouveau modèle de la personnalité en six facteurs et ont également conçu un outil pour les mesurer le HEXACO-PI (*Honesty, Emotionality eXtraversion Agreeableness Conscientiousness Openness to Experience – Personality Inventory*) (Ashton et al., 2004; Lee & Ashton, 2004).

Les dimensions de ce modèle sont

- ⤴ Honnêteté / Humilité
- ⤴ Emotionalité
- ⤴ Extraversion
- ⤴ Agréabilité
- ⤴ Conscience
- ⤴ Ouverture à l'expérience

Ainsi, de Vries (2012) a mis en lien ces dimensions de la personnalité avec quatre styles de leadership : le leadership éthique, le leadership charismatique, le leadership d'encouragement et le leadership centré sur la tâche. Il a observé un lien direct fort entre la dimension Honnêteté / Humilité et le leadership Éthique, l'Extraversion et le leadership Charismatique, l'Agréabilité et le leadership d'Encouragement et enfin la dimension Conscience et le leadership Centré sur la tâche.

4. Les types de personnalité

Dans les années 90, un certain nombre de chercheurs ont adopté une approche de la personnalité conceptuellement différente. En effet, au delà de l'approche de la personnalité en termes de traits telle que nous l'avons vue précédemment avec le modèle de Cattell et celui des *Big Five*, d'autres ont préféré adopter une approche sous forme de types.

Contrairement à l'approche par traits, l'approche par types s'intéresse à la personnalité sous l'angle des préférences. Celles-ci sont par nature dichotomique et non pas continues comme le sont les traits. La théorie des types de personnalité la plus populaire aujourd'hui prend ses racines dans l'approche de Jung (1921) qui propose de différencier les individus en fonction de leur préférences fondamentales sur trois dimensions distinctes.

- ✦ Tout d'abord, les individus se différencient selon leur préférence quant à la source et la direction de leur énergie psychologique.

Cette préférence est dénommée Attitude, ses deux pôles dichotomique sont l'attitude **Introvertie** (énergie orientée vers le monde intérieur) et l'attitude **Extrovertie** (énergie orientée vers le monde extérieur).

- ✦ La seconde source de différences est liée à la préférence quant à la collecte, au recueil d'information, qui peut se faire de deux manières opposé : par les sens (**sensation**) ou par **l'intuition**.
- ✦ La troisième source de différences se trouve dans la préférence individuelle quant au traitement de l'information qui peut être traitée par la **Pensée** ou le **Sentiment**.

Enfin, une quatrième source de différences (mises en lumière par Myers, 1962) concerne la préférence en terme de Style de vie : **Jugement** (un style de vie structuré et organisé) ou **Perception** (style de vie spontané et ouvert).

Dans les années 60, Myers (1962), partant de la théorie de Jung, a développée un outil d'évaluation du type psychologique MBTI (*The Myers-Briggs Type indicator*). Qui est aujourd'hui l'outil d'évaluation des types de personnalité le plus utilisé.

Utilisant le MBTI, Nordvik et Brovold (1998) ont mis en place une recherche particulièrement riche. Ils ont étudié quatre catégories de tâches de leadership (Adizes, 1979, 1987): les PAEI (*Production, Administration, Entreprising et Intégration*). Les activités des leaders dans ces tâches ont ensuite été mises en lien avec les quatre dimensions de personnalité du MBTI Introversiion-Extraversiion; Intuition – Sensation; Sentiment – Pensée et Perception – Jugement.

Afin de rendre les résultats de leur étude comparables avec celles menées à l'aide du modèle des *Big Five*, Nordvik et Brovold (1998) ont associé les 4 dimensions du MBTI aux traits suivants issus des Big Five : respectivement Extraversiion, Ouverture, Agréabilité et Conscience (I. B. Myers & McCaulley, 1985).

Leurs résultats ont montré que le leadership dans le domaine de l'administration de tâches est associé avec un faible niveau d'ouverture et un haut niveau de conscience.

Le leadership dans les activités d'*entreprising* (activités liées à la mise en place de nouveaux projets ou le commencement de nouvelles activités) sont liées avec un haut niveau d'ouverture et un faible niveau de conscience.

Le leadership dans les activités d'intégration sont quant à elles liées à l'agréabilité et celles des activités de production à l'introversiion.

Mais il n'y a pas que la personnalité, l'intelligence émotionnelle à elle aussi été beaucoup étudiée au regard du leadership.

B L'intelligence émotionnelle

L'intelligence émotionnelle englobe la capacité à identifier, utiliser, comprendre et gérer les émotions. Un grand nombre de qualités inter- et intra-personnelles sont associées avec l'intelligence émotionnelle. Ainsi, de nombreuses recherches ont trouvé des liens forts entre l'intelligence émotionnelle et l'efficacité du leaders (Kerr, Gavin, Heaton, & Boyle, 2005; Leban & Zulauf, 2004; Srivastava & Bharamanaikar, 2004), alors que d'autres ont fait état d'une absence de liens (Barbutto & Burbach, 2006; Barchard, 2003; Brown, Bryant, & Reilly, 2006). Enfin, d'autres encore critiquent l'intérêt même du concept d'intelligence émotionnelle dans la mesure où celle-ci posséderait trop de variance commune avec les mesures traditionnelles de personnalité (BeShears, 2005).

En ce qui concerne le leadership transformationnel, beaucoup considèrent l'intelligence émotionnelle comme une composante, ou du moins comme étant une variable prédictive de son apparition ou nécessaire à son efficacité (Ashforth & Humphrey, 1995; Conger & Kanungo, 1987; Fox & Amichai-Hamburger, 2001). Ashkanasy et ses collaborateurs, (2000) décrivent même le leadership transformationnel comme « un ensemble de comportements de gestion des émotions du leader et de ses suiveurs. »

Dans sa méta-analyse, Mills (2009) rappelle que certaines des dimensions du leadership transactionnel entretiennent une grande similarité de nature avec les dimensions de l'intelligence émotionnelle (Tableau W.1) :

Dimensions de Leadership Transformationnel	Éléments du modèle d'Intelligence Émotionnelle			
	Bar-On (2006)	Dulewicz et Higgs (2000)	Mayer et Salovey (1997)	Goleman et al., (2002)
Influence Idéalisée	Intra-personnel	Conscience		Management de soi
Inspiration Motivationnelle	Humeur générale	Motivation	Gestion des émotions	Management de la relation
Simulation Intellectuelle	Adaptabilité	Intuition	Facilitation de la pensée	
Considération Individuelle	Inter-personnel	Sensibilité Interpersonnelle	Perception des émotions	Conscience sociale
	Gestion du stress	Résistance émotionnelle	Compréhension des émotions	Conscience de soi
		Conscience de soi		

Tableau W.1 : Comparaisons des dimensions de leadership transformationnel avec les dimensions d'intelligence émotionnelle selon 4 approches de ce concept.

Issue de Mills L.,B. (2009) A Meta-Analysis of the Relationship Between Emotional Intelligence and Effective Leadership. Journal of Curriculum and Instruction (JoCI). p23

Barling et ses collaborateurs (2000) ont ainsi montré un lien entre trois aspects du leadership transformationnel (influence idéalisée, inspiration motivationnelle et considération individuelle) et les dimensions d'intelligence émotionnelle issues du modèle de Bar-On (1997). De même, Leban et Zulauf (2004) ont montré une corrélation importante entre les dimension de leadership transactionnel (tel que mesuré par le MLQ, Bass & Avolio, 1990) et une mesure composite d'intelligence émotionnelle de Mayer et collaborateurs (2003) MSCEIT (*Mayer-Salovey-Caruso Emotional Intelligence Ability Test*). Mandel et Pherwani (2003) ont quant à eux montré que les scores d'intelligence émotionnelle étaient de bons prédicteurs du style de leadership transformationnel. Se même, dans deux recherches différentes, Caruso (Caruso, Mayer, & Salovey, 2002) et

George (2000) ont montré qu'en ce qui concerne le leadership transformationnel, la reconnaissance des émotions et la réponse adéquate aux émotions des subordonnés étaient liés à une efficacité et à une bonne relation leader-subordonné.

Enfin, suite à sa méta-analyse, Mills (2009) conclut, lui aussi, que l'intelligence émotionnelle peut aujourd'hui être considérée comme une part importante de l'efficacité du leader.

Remettant en lumière la question de l'indépendance de l'intelligence émotionnelle, Rubin et ses collaborateurs (2005) ont eux aussi montré que la reconnaissance des émotions, l'affectivité positive et l'agréabilité prédisaient positivement des comportements de leadership. Cependant, ils ont également montré que l'extraversion tendait à modérer la relation entre la reconnaissance des émotions et le leadership transformationnel.

Enfin, dans une recherche controversée par la suite Prati ses collaborateurs (2003) ont démontré qu'une Intelligence émotionnelle élevée est associée avec des leaders plus charismatiques et plus transformationnels. La controverse reste ouverte, bien que plusieurs autres auteurs (Antonakis & House, 2002; Shamir et al., 1993) ont montré qu'un plus grand charisme et la présence de composantes transformationnelles pouvaient être expliqués sans utiliser le concept d'intelligence émotionnelle.

Une décennie plus tard Cavazotte et ses collaborateurs (2012) ont également montré que bien que l'intelligence émotionnelle semblait être significativement reliée au leadership transformationnel, si les auteurs contrôlaient les dimensions de personnalité et d'aptitude intellectuelle cet effet devenait non-significatif .

Nous voyons donc que l'intelligence émotionnelle reste un concept discuté, bien que son implication dans le leadership – directement ou à travers les traits de personnalité – n'est pas à négliger.

Ainsi, au vu de tout cet ensemble de liens entre le leadership et les mesures de personnalité et d'intelligence émotionnelle, peut on encore parler *du* leadership ?

Nous montrons dans ce chapitre que le leadership est loin d'être un concept unitaire aussi bien au niveau théorique qu'au niveau des liens que cette dimension entretient avec d'autres variables en aval (mesure de l'efficacité) et en amont (caractéristiques des leaders sur des variables de personnalité et sur d'autres variables conatives au sens large). Cette complexité du construit entraîne par conséquent une complexité de la mesure. Ainsi, en passant en revue les effets possiblement attribuables au leadership (les conséquences d'un leadership efficace) nous nous rendons compte que non seulement la mesure du leadership sera forcément plus complexe que les modèles ne laissent penser, mais surtout, que le leadership est de par sa définition et sa fonction même une dimension en relation. Un leader n'est leader qu'en présence de suiveurs, d'employés ou de subordonnés. Nous verrons donc que les outils, tout comme les théories ont souvent gagné, avec le temps, en complexité. Cette complexité se retrouve aussi bien en ce qui concerne l'objet de la mesure, qu'en ce qui concerne la manière dont les comportements de leadership sont mesurés. La partie suivante abordera plus directement ces mesures, les outils développés pour les appréhender et leur spécificité. Enfin, la dernière partie théorique se penchera, quant à elle, sur les manières de mesurer et la place de plus en plus importante donnée à l'hétéro-évaluation dans le cadre des évaluations du leadership.

Chapitre 3. La mesure du leadership

Nous avons vu précédemment l'évolution des définitions du leadership qui, avec le temps, sont devenues de plus en plus complexes et de plus en plus inter-reliées aux autres concepts psychologiques comme la personnalité ou les dimensions conatives. De même, nous nous sommes penchés sur les difficultés de mesurer l'efficacité du leadership.

Aussi, comme la question de la mesure est un élément central de la psychologie différentielle – nous allons à présent nous intéresser à la mesure du leadership à proprement parler.

Tout comme les théories que ces outils ont souvent permis de valider, les questionnaires psychométriques que nous allons aborder sont souvent rattachés à un modèle de leadership spécifique. Néanmoins, ils ont en commun l'interrogation sur la place du leader dans son évaluation. Aussi, bien que nous abordons dans ce chapitre les questions de l'hétéro-évaluation et de l'hétéro-évaluation en 360°, nous ne ferons qu'effleurer ce sujet, car étant donné l'importance de cette question, nous avons choisi d'y consacrer entièrement le chapitre suivant (Chapitre 4).

Dans cette revue des questionnaires psychométriques, nous allons aborder les outils évaluant le style de leadership, basés sur les approches unitaires et comportementales du leadership. Puis nous nous intéresserons aux outils mesurant caractéristiques des modèles charismatiques, transactionnelles ou transformationnelles du leadership. Enfin, nous nous attarderons sur les nouveaux outils de mesure du leadership. Intégrant des variables conatives et de personnalité, ces inventaires proposent de dresser un portrait exhaustif des qualités d'un manager.

La grande variété des outils de mesure du leadership que nous allons présenter est directement proportionnelle à la quantité de modèles théoriques que nous avons détaillé précédemment.

1. La mesure des qualités managériales et de leadership

Dès les années 70 apparaissent des outils visant à évaluer les pratiques et les qualités managériales – ces instruments sont les précurseurs des questionnaires psychométriques de mesure du leadership à proprement parler.

A. Organization Description Questionnaire, House et Rizzo (1972)

Le développement de cet outil a été basé théoriquement sur deux concepts. D'une part la théorie mécanique-organique de l'organisation datant des années 60 (Burns et Stalker, 1961) et positionnant chaque organisation sur un continuum allant d'une structure hautement organique à une structure mécanique de l'organisation, qui en est l'opposée.

La structure organique se caractérise par une organisation peu hiérarchique, préférant les règles implicites et fonctionnant sur un mode personnel. La structure mécanique, quant à elle, est fortement hiérarchisée avec des règles et des rôles explicites et prônant une approche strictement professionnelle de l'activité.

D'autre part la théorie X-Y de McGregor (1960) que nous avons décrite précédemment (Chapitre 1.C.2).

L'ODQ (*Organization Description Questionnaire*) est composé de huit échelles bipolaires :

1. Conflit et inconsistance
2. Formalisme
3. Consensus et clarté de l'objectif à atteindre
4. Pression professionnelle
5. Sélection basée sur les aptitudes et la performance
6. Tolérance de l'erreur
7. Adéquation du planning
8. Actions adéquates de la part de l'autorité.

Les modalités de réponses se présentent sous forme d'une échelle de Lickert en sept points, allant de « très peu » à « beaucoup ».

Pour la validation de leur outil, les auteurs ont adopté la méthode proposée par Campbell et Fiske (1959) et portant sur la nécessité de faire la démonstration d'une part de la

validité convergente et d'autre part de la validité divergente d'un outil afin de le valider.

Ces échelles ont ainsi pu être mises en relation et se sont trouvées fortement reliés à nombre de qualités des leaders tels que la considération du leader, l'accent mis par le leader sur la production, l'initiation de structure de la part du leader, etc.

B. *Managerial Practices Survey* de Yukl, (1990)

Le MaPS (*Managerial Practice Survey*) est un outil développé par Yukl dans les années 90 et est basé sur la théorie de l'action raisonnée (Ajzen & Madden, 1986).

Cet outil se propose d'évaluer le comportement du manager sur 14 différentes dimensions représentant différentes fonctions managériales (Yukl, 1994).

1. Planification / Organisation (formulation d'objectifs à long terme, la mise à disposition de ressources nécessaires...)
2. Résolution de problèmes (reconnaissance des problèmes dans le travail, leur analyse et résolution efficace)
3. Monitoring (vérification de la qualité et de la quantité de travail de chacun, évaluation de l'équipe)
4. *Networking* – travail du réseau (développement de contact avec des personnes potentiellement intéressantes pour l'entreprise, maintien des contacts dans la durée...)
5. Informer (donner les informations nécessaires sur les décisions prises, les plans, les changements..)
6. Clarifier (donner des tâches claires, préciser les rôles de chacun, les étapes nécessaires à la réalisation d'un travail...)
7. Motiver et Inspirer (utiliser des techniques d'influence, basées sur la logique et ou les émotions pour générer de l'enthousiasme et de l'envie dans la réalisation des tâches..)
8. Gestion de conflit et team-building (faciliter la résolution pacifique des conflits et maintien de bonnes relations au sein de l'équipe)
9. Soutenir (agir de manière agréable et sensible, être patient et aidant envers les membres de l'équipe)
10. Consulter (discuter avec les autres avant de prendre des décisions et de mettre en place des changements les concernant)

11. Reconnaissance (faire preuve de reconnaissance et de valorisation envers les réussites et les bonnes performances de chacun.)
12. Développement (proposer des pistes d'évolution professionnelle et faciliter le développement de compétences tout au long de la carrière.)
13. Récompense (proposer des récompenses suffisantes comme des hausses de salaire ou des promotions lorsque des niveaux de performance sont atteints)
14. Déléguer (permettre aux autres d'exercer des responsabilités et leur donner le pouvoir de prendre certaines décisions)

Dans le questionnaire, le sujet doit répondre sur une échelle de type Lickert en quatre points.

Un grand nombre d'études ont été menées pour valider cet outil et surtout éprouver sa validité de critère. Ces études réalisées auprès de gestionnaires de salons de beauté, de cadets militaires, de managers des ventes et de directeurs d'écoles élémentaires (Yukl et al., 1990) ont présenté des résultats concluants. Pourtant, au début des années 2000, d'autres instruments ont été préférés par Yukl et ses collègues au MaPS, qui conseillent de réserver ce dernier pour une utilisation dans le cadre de programmes de développement managérial et sous forme de questionnaire à 360° (Yukl, Gordon, & Taber, 2002).

2. Les mesures du type et du style de leader

Lors de notre étude des théories du leadership, nous avons vu l'évolution des théories et des définitions du leadership, s'intéressant aux approches personalistes et situationnelles puis à la notion de style. Nous allons voir à présent les outils développés par les chercheurs tenant des modèles de style de leadership.

A. Le Leadership Behaviour Description Questionnaire

La recherche ayant amené au LBDQ (*Leadership Behaviour Description Questionnaire*) a été initiée dans les années 50 par les travaux de Shartle, Hemphill et Coons (Hemphill & Coons, 1957) et menée par les équipes de l'Université de l'Ohio.

Les dimensions du LBDQ correspondent aux dimensions théoriques mises en avant par les chercheurs de l'université de l'Ohio : initiation de structure et considération.

Le LBDQ (Halpin, 1957) est composé de 40 items décrivant chacun un type de comportement possible d'un leader. Les répondants doivent indiquer la fréquence à laquelle ils ont perçu chacun de ces comportements de la part de leur leader. Il s'agit donc d'une évaluation par les subordonnés (la question méthodologique sous-jacente à l'hétéro-évaluation sera abordée dans le chapitre qui suit).

Pour chaque dimension, les réponses (fréquences) de chaque observateur sont moyennées et la somme de ces réponses donne le style préférentiel du leader.

Sur les 40 items présentés, 30 sont effectivement côtés : 15 d'entre eux correspondent à des comportements « d'Initiation de structure » et 15 à des comportements de « Considération ». La consistance des deux échelles est très bonne .83 et .92 respectivement (Halpin, 1957). Ces échelles présentent par ailleurs une corrélation de .38.

De nombreuses études ont utilisé cet outil et ont retrouvé la présence de ces deux dimensions : orientation vers le travail ou vers les employés, et ce dans un grand nombre d'échantillons de provenance différente (dans l'armée, dans l'industrie, l'éducation... Fleishman, 1953a, 1953b).

Dès 1957 le questionnaire est présenté en plusieurs versions : la version d'auto-évaluation pour le manager et la version hétéro-évaluation pour ses observateurs, mais aussi une version permettant de décrire le leader idéal et ce aussi bien pour le manager que pour ses collaborateurs. L'idée de cette démarche étant profondément ancrée dans le modèle théorique des chercheurs de l'Université de l'Ohio, à savoir que l'efficacité réside dans l'appariement du style du leader et des demandes de la situation.

Le LBDQ a été actualisé en 1962 (LBDQ *Form -XII*, Stogdill, 1962) dans le cadre de ce travail de nouvelles dimensions expérimentales ont été rajoutées au questionnaire (dix) et le nombre de questions est passé à 100. Stogdill (1962, 1969) a éprouvé cette nouvelle version du questionnaire à travers des validations à base expérimentale auprès de militaires, de managers des autoroutes et de leaders communautaires. La nouvelle version présente d'excellentes qualités de consistance interne et de validité de critère qui confirment les résultats précédents.

Pourtant ce questionnaire présente des limites. En effet, il n'existe pas de normes pour le LBDQ car celui-ci a été construit pour être utilisé dans le cadre de recherches. Aussi, les auteurs déconseillent fortement de l'utiliser dans le cadre de sélections ou d'évaluations (Stogdill, 1962). D'ailleurs, bien que le LBDQ ait connu quatre révisions, il n'a plus été révisé depuis les années 70.

B. Le Leader Opinion Questionnaire

Le LOQ (*Leader Opinion Questionnaire*) de Fleishman (Fleishman, 1960; Fleishman & Peter, 1962) s'inscrit lui aussi, dans la grande famille des outils de l'Université de l'Ohio. À la différence de la dernière version du LBDQ qui propose un ensemble de sous-échelles en complément des deux dimensions principales, le LOQ permet, lui, d'évaluer seulement les dimensions de considération et d'initiation de structure.

La dimension considération définit la mesure dans laquelle un leader recherchera, au travail, des relations interpersonnelles basées sur la confiance mutuelle, le respect des idées des subordonnés et la prise en compte de leurs sentiments.

Ainsi un score élevé sera associé à un bon climat de travail et à une communication bi-directionnelle. Un score faible sera quant à lui descriptif d'un comportement plus impersonnel de la part du manager.

La dimension Initiation de structure définit la mesure dans laquelle un leader sera à même de spécifier clairement son rôle et celui de ses subordonnés dans l'atteinte de leur objectif. Ainsi un score élevé sur cette dimension décrira des individus plus directifs dans les activités de groupe, à travers la planification, l'organisation et la communication.

Le LOQ est composé de 40 propositions rattachées soit à des comportements de considération, soit à des comportements d'initiation de structure à évaluer sur une échelle allant de « Jamais » à « Toujours ». Tous les items sont construits sous forme de situations pour lesquelles l'évalué doit se positionner selon sa perception du comportement à adopter. Aussi, contrairement au LBDQ, cet outil est conseillé et utilisé dans le cadre de recrutements et de processus de sélection.

Pourtant, au vu de la complexité des théories actuelles du leadership, cet outil présente un intérêt limité car il ne s'intéresse qu'aux deux dimensions centrales du modèle de l'Université de l'Ohio. Nous allons à présent nous pencher sur un questionnaire lié au modèle de Blake et Mouton, qui avaient proposé au niveau théorique de réunir ces deux dimensions (Initiation de structure et Considération) pour créer cinq styles particuliers de leadership.

C. Managerial Grid Questionnaire

Faisant suite à leur théorie basée sur la grille managériale et définissant cinq styles de leadership, le *Managerial Grid Questionnaire* (Blake & Mouton, 1978) vise à évaluer le style de leadership d'un manager donné.

Le style de leadership des répondants est déterminé grâce à un processus en deux temps.

Tout d'abord, pour évaluer la communication dans le processus managérial, le manager doit classer cinq paragraphes. Chacun des paragraphes correspond à l'un des cinq styles de leadership issus de la grille managériale. Les candidats doivent donc ordonner ces cinq styles en fonction de leur similarité avec leur propre fonctionnement.

Dans un second temps, six éléments de style managérial sont présentés (décision, conviction, conflit, humeur, humour et effort) avec cinq alternatives de réponse

(correspondant toujours aux cinq styles de la grille managériale). Les répondants doivent sélectionner la réponse – et donc le style – qui les décrit le mieux dans chacune des situations. Sur la base de ces deux étapes, le style de leadership du manager est déterminé en sélectionnant le style le plus souvent choisi par le répondant.

Le *Managerial Grid Questionnaire* est principalement utilisé pour permettre au manager une prise de conscience de son propre style. Il peut donc être utilisé en amont de formations ou de sessions de développement managérial.

S'agissant d'un classement des styles réalisé par le répondant, concernant lui-même, il n'y a donc pas de normes pour ce questionnaire – et les résultats sont par conséquent difficilement comparables d'un individu à un autre. Enfin, cet outil a surtout été utilisé par les auteurs pour éprouver et tester leur modèle, mais il est peu utilisé depuis.

D. Leadership Adaptability and Style Inventory (LASI)

Tout comme Blake et Mouton (1978) Hersey et Blanchard ont proposé un modèle de l'efficacité du leader basé non seulement sur son adéquation avec la situation mais tenant également compte du niveau de développement des subordonnés.

Comme nous l'avons vu dans le premier chapitre (Chapitre 1.2.C), selon la combinaison situation / type de subordonnés, différentes orientations du leader vont s'avérer efficaces.

Hershey et Blanchard présentent la première version du LASI en 1974 (Hersey & Blanchard, 1974) – renommé ensuite *Leadership Effectiveness and Adaptability Description*. Ce questionnaire est composé de 12 situations pour chacune desquelles le manager doit choisir une des quatre alternatives proposées. Les quatre choix de réponse correspondent à l'un des quadrants de la théorie de Hershey et Blanchard (Forte Tâche et Forte Relation; Forte Tâche et Faible Relation; Faible Tâche et Forte Relation et enfin Faible Tâche et Faible Relation)

Le test se propose de mesurer trois composantes de la perception du leader de son comportement: son style, l'étendue de son style et l'adaptabilité de son style.

Le style est défini comme un ensemble de comportements utilisés pour influencer l'activité d'autres personnes. Ces comportements sont orientés soit vers la tâche (organisation, planification..) ou vers la relation avec les autres (communiquer avec les autres, prendre

en compte leurs attentes, les responsabiliser..)

Pour déterminer son style, le manager doit comptabiliser le nombre de réponses « Tâche » et le nombre de réponses « Relations » sur l'ensemble des 12 situations.

Dès 1974, les auteurs notent une grande différence entre l'auto attribution (perception) d'un style et son style effectif, ils proposent donc une version Observateurs pour le LASI. De la sorte ce ne sont plus seulement les leaders qui s'évaluent eux-même, mais aussi leur subordonnés ou collaborateurs qui les évaluent également. Nous aborderons plus en détail les spécificités de cette approche en 360° dans la partie suivante.

L'étendue des styles s'analyse en comparant la fréquence respective de chacun des quatre styles chez une même personne.

Enfin l'adaptabilité du style (également appelé efficacité du style) correspond à l'adéquation du style à la situation. Il se mesure en associant à chaque modalité de réponse un score allant de moins deux à plus deux caractérisant l'efficacité de ce comportement spécifique dans la situation présentée.

Ainsi, en ajoutant la dimension efficacité, les auteurs proposent une mesure en accord avec le modèle Tri-Dimensionnel de l'efficacité du leader (Reddin, 1986).

Le questionnaire et le modèle sous-jacent ont été validés par de nombreuses recherches. Dans le cadre d'une démarche expérimentale, Hersey et Blanchard (1982) ont pu mettre en lien la maturité des individus avec l'efficacité d'un style de leadership donné auprès de participants à un programme de formation managériale. Toujours en utilisant des paradigmes expérimentaux, plusieurs recherches ont montré des liens entre style de leadership évalué par le LASI et efficacité. Ainsi une corrélation positive a été trouvée entre les auto-évaluations par les managers de l'adaptabilité de leur style de management et l'évaluation par les subordonnés de l'efficacité de leurs équipes (Graeff, 1997; Jacobsen, 1984). De même, la flexibilité rapportée par des femmes manager avec le LEAD s'est trouvée corrélée avec leur niveau d'efficacité (Kohut, 1983).

Pourtant, ces résultats ne font pas l'unanimité. En ce qui concerne les propriétés métriques fondamentales du LASI, des doutes subsistent sur sa consistance interne et sur sa capacité à être utilisé dans tous les types d'environnements (Graeff, 1997). De fait, les relations entre flexibilité et efficacité n'ont été retrouvées ni auprès de gestionnaires

d'unités de tissage (Narayanan, Venkatachalam, & Bharathiar, 1982) ni auprès des managers de chaînes de magasins (Goodson, McGee, & Cashman, 1989).

Aussi, du fait du manque de consistance interne de l'outil (Brief & Aldag, 1981) et des résultats mitigés obtenus, ce questionnaire n'est aujourd'hui plus très utilisé.

E. *The least preferred Co-worker*

Fiedler (1978) auteur du modèle de contingence du leadership a développé un outil permettant d'évaluer les dimensions qui y sont associées. Pour ce faire il propose d'utiliser l'échelle LPC (*Fiedler's Least Preferred Coworker*) qui mesure les réactions affectives du leader envers son employé le moins apprécié.

Le questionnaire LPC est composé d'une série d'adjectifs bipolaires. Le manager doit décrire son subordonné le moins apprécié en le positionnant sur une échelle en huit points entre deux extrémités, telles que « aidant - frustrant », « intéressant - ennuyeux »... Le pôle positif de chaque paire d'adjectifs donne le maximum de points, le score au LPC correspond à la somme des points sur l'ensemble des items.

Ainsi, un professionnel ayant un score élevé est considéré comme étant capable d'associer à son subordonné le moins apprécié des traits aussi bien positifs que négatifs, alors qu'un professionnel avec un faible score de LPC voit ce collaborateur en termes uniquement négatifs.

Concernant les qualités métriques, la première version de l'échelle présentait des problèmes de consistance internes qui ont obligé Fiedler (1978) à effectuer une révision de l'échelle afin de minimiser le facteur orientation vers la tâche et proposer 18 nouveaux items. Cette révision a permis d'améliorer sensiblement la consistance interne de .79 à .91 selon les études (Rice, 1978 a).

De même Rice (1978b) dans sa méta-analyse, trouve une bonne stabilité dans le temps pour les résultats à ce questionnaire (valeur médiane à .6 pour des écart allant de quelques semaines à plus de 50 semaines) si le même subordonné est décrit à chaque fois.

Pourtant beaucoup de controverses ont eu lieu quant au sens du score de cette échelle. Ainsi, à l'origine, les leaders avec un faible LPC étaient simplement considérés comme orientés vers le travail et inversement, ceux avec un score élevé, étaient considérés comme orientés vers le relationnel (Fiedler, 1978). Quelques années plus tard, les mêmes auteurs ont proposé une interprétation en terme de distance sociale (Fiedler, 1967). Enfin, dans les études plus récentes, les scores au LPC ont également été interprétés en terme de complexité cognitive (Foa, Mitchell, & Fiedler, 1971; Hoffman & Roman, 1984; Vecchio, 1979).

3. Les mesures multi-factorielles du leadership Transformationnel et Charismatique

Avec les questionnaires précédents, les outils proposés par les tenants des approches du leadership en termes de style et de type ont été étudiés. Nous allons maintenant passer à une conception quelque peu différente du leadership basées sur l'approche de Weber et construite autour du concept du leader charismatique. Nous allons donc aborder les outils se proposant d'évaluer les dimensions propres aux leaders charismatiques et transformationnels.

A. Le Leadership Charismatique

1. C-R Scale of Charismatique Leadership

Conger et Kanungo (1997) proposent un outil pour mesurer le leadership charismatique, basé sur leur modèle. Or, le modèle du leader charismatique de Conger et Kanungo (1994) est par essence un modèle par attribution. Les auteurs définissent donc un leader en tant que charismatique s'il est perçu par ses suiveurs comme ayant la capacité de leur inspirer une vision et de mettre en place des actions qui donneraient une dimension « extraordinaire » à la mission de tous, en un mot s'il est perçu comme charismatique.

La mesure de ces traits doit donc obligatoirement se réaliser à travers la perception des suiveurs, car étant une théorie par attribution, les subordonnés sont les seuls aptes à juger de « l'extraordinarité » de leur leader.

Il s'agit donc d'un outil proposant une mesure sous forme d'hétéro-évaluation. Il est important de ne pas confondre ce cas particulier avec les approches à 360 degrés que nous détaillerons dans le chapitre suivant (Chapitre 4). En effet, le test *C-R Scale of Charismatique Leadership* utilise l'évaluation des subordonnés pour établir si le leader est ou non un leader charismatique, mais il ne propose pas à d'autres personnes d'évaluer ce même leader ni au leader de s'auto-évaluer, ce que fait une approche par 360°.

La première version de l'échelle était composée de 25 items et mesurait six dimensions :

1. Sensibilité à l'environnement
2. Sensibilité aux besoins des subordonnés
3. Comportement in conventionnel
4. Vision stratégique et articulation de celle-ci
5. Prise de risque
6. Rejet du *statu quo*

Cette version présentait de bonnes qualités métriques (bonne stabilité dans le temps et validité convergente et divergente satisfaisante) et une structure factorielle suffisamment stable (Conger & Kanungo, 1994).

Pourtant, un examen plus approfondi des items a pu montrer une certaine dose de redondance et une tendance de certaines questions à saturer sur plusieurs facteurs (Conger & Kanungo, 1994). Aussi une version révisée de l'échelle CKS (*Revised C-K scale of Charismatic Leadership*) à été proposée.

Lors de cette révision le sixième facteur (Rejet du *statu quo*) qui n'était composé que de deux items, n'a pas été retrouvé lors d'étude de la structure du questionnaire. Aussi, les auteurs ont proposé un questionnaire composée de 20 items à noter par les subordonnés et mesurant cinq dimensions.

La structure de ce nouveau questionnaire a été éprouvée lors d'études de validation auprès de gestionnaires américains, de délégués de partis politiques canadiens et d'employés de bureau indiens. Ces trois grandes études ont permis de démontrer une bonne consistance interne et une stabilité des résultats mais elles ont surtout permis d'examiner la validité concurrente et discriminante de l'échelle.

Cependant, en ce qui concerne les éléments de validité convergente, les dimensions de leadership charismatique de l'échelle de Conger et Kanungo présentent des liens positifs mais d'amplitude variée (de modéré à fort) avec les dimensions de leadership transactionnel dont ils sont conceptuellement très proches (corrélation entre .27 et .72 selon les études – (Conger & Kanungo, 1997).

De même, bien que les dimensions du leadership transactionnel ne sont pas censées être reliés au leadership charismatique, des corrélations positives ont pu être observées (Rowold & Kersting, 2008). En ce qui concerne les autres modèles du leadership, des

liens parfois contradictoires ont été observés d'études en études entre les dimensions du questionnaire de Conger et Kanungo et les dimensions Considération (corrélations entre .03 et .14) mais surtout d'Initiation de structure (-.2 à .13) (Rowold & Kersting, 2008).

Enfin, le questionnaire étant particulièrement court, et certaines dimensions n'étant composées que de trois items, des doutes peuvent persister sur la consistance inter-juges (dimension non abordée dans les études de validation).

La dernière remarque concernant ce questionnaire – et expliquant peut-être sa faible utilisation – est son ancrage dans la théorie de l'attribution. En effet cette théorie nécessite obligatoirement une évaluation par les subordonnés et ne permet pas au manager de s'auto-évaluer, ce qui, au vu des théories actuelles est une grande limite pratique.

2. Ethical Leadership Scale (ELS)

L'idée de mesurer la dimension éthique du leadership est bien sûr associée – en ce qui concerne la théorie – aux approches de Burns (1978). Cela dit, pour de nombreux autres auteurs (Howell & Avolio, 1992; Trevino & Weaver, 2003) la dimension éthique s'inscrit totalement dans les définitions transformationnelles des leaders.

Ainsi, pour les auteurs de l'ELS (*Ethical Leadership Scale* - (Brown, Treviño, & Harrison, 2005) la dimension éthique représenterait une petite composante qui s'inscrirait au niveau des dimensions « Inspirer une vision » et « Stimuler les autres » – qui sont toutes deux des caractéristiques des leaders Charismatiques et Transformationnels.

Dans l'ELS, Brown et ses collègues (2005) définissent le leadership éthique comme « la démonstration d'une conduite normative à travers ses actions personnelles et ses relations inter-personnelles. De plus, il s'agit également de la promotion de cette conduite auprès de ses subordonnés par une communication bi-directionnelle, le renforcement et la prise de décision » (Notre traduction Brown et al., 2005, p. 120⁴).

Partant de cette définition, L'ELS propose de mesurer la perception des employés des comportements éthiques de leurs supérieur – que ceux si soient des leaders formels ou

⁴ We define ethical leadership here as the demonstration of normatively appropriate conduct through personal actions and interpersonal relationships, and the promotion of such conduct to followers through two-way communication, reinforcement, and decision-making.

informels. Il s'agit donc comme pour le CKS d'une échelle d'hétéro-évaluation.

Ce questionnaire unidimensionnel est composée de dix items à coter par les subordonnés sur une échelle de type Lickert en cinq points allant de « Pas du tout d'accord » à « Tout à fait d'accord ».

Les études de validation ont pu montrer que cet instrument présentait une bonne consistance interne (Alpha de Cronbach = .89) et une bonne stabilité dans le temps (Ponnu & Tennakoon, 2009). En ce qui concerne la validité convergente et divergente, l'ELS se différencie des évaluations de désirabilité sociale et d'influence idéalisée. De même, en envisageant les comportements éthiques comme issus d'un apprentissage social, les auteurs ont montré que les leaders perçus comme éthiques s'engageaient plus souvent dans des comportements d'aide et présentaient moins de comportements non-éthiques (mensonge, vol, sabotage) (Brown et al., 2005). Ce concept et son évaluation semblent donc apporter un élément complémentaire aux évaluations du leadership transactionnel tel que réalisé avec des outils classiques.

Pourtant l'échelle reste très limitée, étant ciblée sur une dimension seulement, elle est principalement utilisée en complément d'autres outils plus larges, mesurant les dimensions du leadership à proprement parler.

B. Le leadership Transformationnel

1. Multifactor Leadership Questionnaire (MLQ)

Le *Multifactor Leadership Questionnaire* (Bass & Avolio, 1990a, 1990b) est directement lié à la théorie de leader Transactionnel et Transformationnel de Bass.

En effet, Bass conçoit le leadership comme un continuum entre le domaine du non leadership (laissez-faire) et le leadership transformationnel (basé sur le charisme existant et attribué) en passant par le leadership transactionnel, qui repose lui sur l'application de punitions et récompenses.

Depuis sa création dans les années 90, le questionnaire a subi plusieurs révisions. La version actuelle comporte 50 items.

Neuf échelles sont évaluées (cinq portent sur le leadership transformationnel, trois sur le leadership transactionnel et une sur le laissez-faire).

Le **leadership Transformationnel** est mesuré grâce à cinq dimensions ;

- ✦ Le charisme attribué (huit items) correspond à la dimension charisme à proprement parler.
- ✦ L'influence idéalisée (dix items) correspond aux attributions faites par les subordonnés et se base sur la confiance en soi et le succès projeté par le leader.
- ✦ Le leadership inspiré (dix items) est lié aux comportements où le leader articule et partage les objectifs, en instaurant une bonne compréhension mutuelle.
- ✦ La stimulation intellectuelle (dix items) correspond au soutien et à l'encouragement dont fait preuve le leader par rapport aux idées de ses subalternes.
- ✦ Considération individuelle (huit items) correspond à l'attention que le leader accorde à chacun en fonction de ses besoins.

Le **leadership Transactionnel** est mesuré grâce à trois facteurs. Ce type de leadership correspond à une démarche du leader qui vise à obtenir l'accord des subordonnés pour accomplir les tâches requises :

- ✦ La récompense offerte (neuf items) – le leader emploie la récompense pour obtenir la réalisation de ces tâches (renforcement positif des comportements permettant d'atteindre les objectifs)
- ✦ Gestion active et passive par exception (sept items chacun) – le leader emploie la coercition et la punition pour obtenir la réalisation de ces tâches :
 - Management Actif par Exception: type de management où le leader recherche activement l'erreur de ses subordonnés
 - Management Passif par Exception : type de leader qui attendent passivement d'apprendre l'existence des erreurs pour prendre des mesures.

Enfin, le **leadership laissez-faire**, est assimilé à une absence de leadership et il est évalué grâce à une seule dimension composée de huit items.

Trois échelles supplémentaires proposent d'évaluer certains des effets du Leadership (échelles abordées dans la partie précédente – Chapitre 2.1.B et Chapitre 2.1.C) : il s'agit d'une variable d'efforts supplémentaires, d'une mesure d'efficacité et d'une échelle de satisfaction.

Tous les items sont à coter sur une échelle de type Lickert en cinq points allant de « Fréquemment si ce n'est toujours » à « Pas du tout ». Le MLQ existe en version Self (MLQ-5X-Self; Bass & Avolio, 1990a) et observateur (MLQ-5X-Rater; Bass & Avolio, 1990b) ce qui permet au questionnaire d'être administré dans le cadre d'études à 360° dont nous reparlerons dans le chapitre suivant (Chapitre 4).

Outil très utilisé et très étudié, le MLQ possède une très bonne consistance interne (Totale: .83; Partie Leadership Transformationnel = .93; Partie Leadership Transactionnel = .56). Le test a également montré des liens forts entre les scores aux dimensions transactionnelles et certaines variables d'effet telles que l'efficacité, la satisfaction et l'effort supplémentaire fourni (Gasper, 1992; Lowe, Kroeck, & Sivasubramaniam, 1996; Patterson, Fuller, Kester, & Stringer, 1995) comme cela a pu être vu dans le chapitre 2 (Chapitre 2.1).

Pourtant, un ensemble d'études menées en dehors des États-Unis vers la fin des années 90 se sont, elles-aussi, penchées sur ce questionnaire (Carless, 1998; Den Hartog, Van Muijen, & Koopman, 1997) et leurs résultats se sont avérés moins évidents. Ainsi, dans les deux études menées aux Pays Bas et en Australie, les auteurs sont parvenus à retrouver les trois structures correspondant aux trois types de leadership (transformationnel, transactionnel et laissez-faire) mais ils n'ont pas trouvé les dimensions sous-jacentes correspondantes.

De même, quelques années plus tard, des études menées sur une population américaine ont, elles aussi, remis en cause la structure du MLQ (Northouse, 2007; Tejeda, Scandura, & Pillai, 2001). Dans ces études, les chercheurs ne sont pas non plus parvenus à retrouver « la structure des facteurs de premier ordre » (Tejeda et al., 2001, p. 36). Il semblerait donc que le MLQ présente des résultats controversés quant à sa structure, bien que tous s'accordent sur les trois dimensions globales : Leadership Transactionnel, Leadership Transformationnel et Leadership Passif (laissez-faire).

2. Transformational leadership questionnaire (TLQ)

Développé par Alimo-Metcalfe et Alban-Metcalfe (2001) au début des années 2000, ce questionnaire vise en premier lieu à pallier les limites du MLQ en ce qui concerne son fondement principalement américain mais aussi à proposer une mesure plus spécifiquement centrée sur le leadership transformationnel. Les auteurs cherchent donc à améliorer la différenciation des facteurs sous-jacents au leadership transactionnel.

Travaillant sur le management proche et distant, les auteurs remarquent le peu d'importance que le MLQ accordé à cette variable et considèrent que cela pourrait être un des éléments pouvant expliquer le peu de différences observées chez les managers pour les sous-dimensions du MLQ.

En testant leur outil auprès de leaders de gouvernements locaux, Alimo-Metcalfe et Alban-Metcalfe (2001) ont proposé une première version de leur questionnaire : le TLQ-LGV (*Transformational Leadership Questionnaire - Local Government Version*).

Le questionnaire est composé de neuf dimensions pouvant être organisées en deux groupes de second ordre, tous les deux de type Transactionnel : Orientation – Interne et Orientation – Externe.

Les neuf dimensions sont évaluées sur une échelle de type Lickert en six points allant de « Tout à fait d'accord » à « Tout à fait en désaccord », deux options de réponses complémentaires étaient proposées : « Ne sais pas » et « Ne s'applique pas ».

Ces neuf dimensions sont dénommées comme suit:

- Intérêt sincère pour les autres (*Genuine concern for others*),
- Sensibilité politique et compétence (*Political sensitivity and skills*),
- Détermination et confiance en soi (*Descisivness, determination and self-confidence*),
- Intégrité, honnêteté et ouverture (*Integrity, trustworthy, honest and open*),
- Empowerment et développement de potentiel (*Empowers, develops potentialities*)
- Visionnaire et réseuteur (*Inspirational networker and promoter*)

- Facile d'accès (*Accessible, approachable*)
- Participatif et définit les limites (*Clarifies boundaries, involves others in decisions*)
- Encourage la pensée stratégique (*Encourages critical and strategic thinking*)

Dans leur étude de validation, les auteurs ont établi une bonne consistance interne (Alphas de Cronbach allant de .98 à .97 selon les échelles) et une validité de critère convergente et divergente satisfaisante (corrélations allant de .42 à .49 selon les critères et les facteurs) (Alimo-Metcalfe et Alban-Metcalfe, 2001).

Tout d'abord il est à noter que bien que les auteurs trouvent deux facteurs de second ordre – tous les deux transformationnels – les dimensions sous-jacentes présentent dans le TLQ-LGV des liens parfois forts, d'autres études seraient donc nécessaires pour éprouver cette structure. De même, tout comme le MLQ avait été développé dans un environnement américain, le TLQ-LGV a lui, été construit sur la base de données issus du secteur public du Royaume Uni, il doit donc prouver son utilité et sa structure auprès d'échantillons de population plus larges.

3. Leadership Practice Inventory

Le LPI (*Leadership Practice Inventory*) a été développé par Kouzes et Posner dans les années 90 à partir d'un grand nombre d'études de cas. Les auteurs, visant à promouvoir le développement des qualités de leadership dans les organisations, ont réalisé de très nombreux entretiens auprès de leaders influents et exerçant à des niveaux hiérarchiques très élevés. Ces entretiens leur ont permis de définir 5 grandes dimensions qui seraient, d'après eux, à la base des comportements et actions des leaders efficaces (Kouzes & Posner, 1987, 1993).

- ✧ Remettre en cause les processus (*Challenge the process*).

Cette dimension englobe l'idée de remettre en cause les statuts quo, expérimenter de nouvelles choses, prendre des risques.

- ✧ Inspirer une vision partagée (*Inspiring a shared vision*).

Cette dimension décrit la capacité du leader à proposer une vision attirante de l'avenir et d'attirer les autres à partager sa vision.

- ✧ Permettre aux autres d'agir (*Enabling others to act*).

Le leader doit aider ses subordonnées à se mettre en action, faire preuve d'empowerment.

✧ Montrer la voie (*Modeling the way*)

Le leader doit, de par ses propres actions et valeurs être un exemple pour les autres.

✧ Encourager le cœur des hommes (*Encouraging the heart*).

Enfin, le leader se doit de montrer du feedback, de l'écoute et célébrer les réussites de l'équipe.

Ainsi, pour les auteurs, ces cinq actions et comportements seraient les dimensions permettant de décrire les axes par lesquels se matérialise le leadership transformationnel. Ces axes correspondraient donc aux actions à mettre en place pour réaliser et développer un leadership transformationnel.

Dans le LPI, chacun des facteurs transformationnel est mesuré grâce à six items (30 items au total). Pour chacun des items, le répondant doit se positionner sur une échelle de type Lickert en cinq points allant de « Rarement, voire jamais » à « Fréquemment, toujours »

Un grand nombre d'études ont été menées sur les qualités métriques du LPI (Kouzes & Posner, 1993) et ont présenté des résultats satisfaisants.

Pourtant, dans les études récentes des qualités du test avec la méthode de réponse à l'item IRT (IRT – *Item Response Theory*) Zagorsek et ses collaborateurs (2006) n'ont pas retrouvé de résultats convaincants.

Aussi, Zagoresk et ses collaborateurs préconisent l'utilisation de cet inventaire dans le cadre de sessions de développement personnel et de formations plutôt que dans le cadre de sélection.

Il est par ailleurs à noter que Kouzes et Posner (1987) ont dès le début, principalement mis en avant leur outil comme instrument diagnostique à utiliser en amont de sessions de formations destinées à accroître l'efficacité des managers en les aidant à mettre en place des comportements plus transactionnels.

Pourtant, bien qu'une distinction soit souvent faite entre outils à visée développementale et outils à visée de sélection ou de recrutement, la question de la pertinence de cette distinction se doit d'être posée.

En effet, un outil utilisé dans le cadre de formations vise à aider un individu donné à mettre en lumière ses zones de difficulté tout en identifiant ses forces. Or pour ce faire, l'outil se doit de posséder des qualités métriques de fidélité mais aussi de validité – sinon tout ce travail d'identification ne pourra avoir lieu de manière tangible.

De plus, utilisé par le sujet lui-même, mais aussi par des observateurs externes, les erreurs ne vont faire que « s'additionner » pour, au final, donner une évaluation difficilement compréhensible et ne permettant au sujet d'obtenir qu'un rendu entaché d'erreurs.

La seule situation où cette distinction est pertinente pour nous est le cas d'outils qui, de par leur structure, proposent au sujet de classer un ensemble d'éléments entre eux pour obtenir une hiérarchie personnelle, comme c'est le cas du Managerial Grid Questionnaire qui classe les styles de leadership préférés du sujet (Chapitre 3.2.B). Ces questionnaires, du fait de leur méthodologie permettent uniquement de comparer des classements Intra-Individuels et rendent donc difficile la comparaison de différents candidats entre eux.

4. Global Transactional Leadership Scale (GTL)

Ayant noté l'intérêt croissant pour les mesures du leadership et en même temps la complexité des outils disponibles (MLQ; CKS et LPI) Careless et ses collaborateurs (Careless et al., 2000) ont décidé de construire un outil en version courte pour l'évaluation du leadership transformationnel.

Pour décrire le leaders transformationnel, ils partent de la synthèse proposée par Podsakoff et ses collaborateurs, en 1990.

Dans le GLT (*Global Transactional Leadership Scale*) les auteurs définissent le leader transformationnel à travers sept dimensions:

1. il communique une vision,
2. il aide ses subordonnés à se développer,
3. il les soutient,
4. il développe la prise de pouvoir (*empowerment*),
5. il est innovant voir in-conventionnel,
6. il mène par l'exemple,
7. il est charismatique.

Sur la base de ces dimensions, les auteurs construisent sept items (un par dimension) qui composent le GLT. Cette version courte a montré des liens importants avec les mesures traditionnelles du leadership transformationnel de même qu'avec des variables d'efficacité

du leader (Carless et al., 2000).

Pourtant, de part le nombre d'items par dimension, cette échelle ne peut être envisagée que comme un construit unidimensionnel, et de ce fait, présente les limites de ses qualités. Le GLE est rapide de passation, mais ne permet au final qu'un diagnostic de présence forte, modérée ou faible des caractéristiques de leadership transformationnel chez un manager donné.

5. Empowering Leadership Questionnaire

L'Empowering Leadership Questionnaire (Arnold, Arad, Rhoades, & Drasgow, 2000) a été développé pour répondre au manque de spécificité des outils de leadership transformationnel. En effet, vers la fin des années 90, le concept *d'empowerment* s'est rapidement répandu et la mise en place de techniques *d'empowerment* dans les entreprises est devenue courante.

L'empowerment, terme anglais, correspond aux « notions d'habilitation et de responsabilisation ; ce terme met également en évidence le fait qu'il s'agit de mettre en valeur les contributions et les ressources des collaborateurs. » (Bernaud & Lemoine, 2000, p. 147)

Aussi, les outils existant ne pouvant s'adapter aux spécificités des environnements utilisant les techniques *d'empowerment*, Arnold et ses collaborateurs (2000) décident de proposer un nouvel outil de mesure du leadership.

Sur la base d'entretiens et d'études des situations critiques de management, les auteurs ont défini un premier set de huit axes d'actions.

Les modèles structurels ont finalement orienté le choix d'une organisation de test en cinq dimensions :

1. Mener par l'exemple
2. Former (*Coach*)
3. Prendre des décisions de manière participative
4. Informer
5. Montrer de la considération / interagir

Des liens significatifs ont été montrés entre ces cinq dimensions et les mesures du MaPS (Yukl et al., 1990) pour les fonctions de leader et du LBDQ (Stogdill, 1962) pour les dimensions Initiation de Structure et Considération. Pourtant, conçu pour être utilisé dans des environnements ayant recours aux techniques *d'empowerment*, ni ce questionnaire ni surtout sa structure n'ont été éprouvés en dehors.

Au vu du nombre de questionnaires que nous avons abordé, nous pouvons aisément nous rendre compte de la place que le modèle de leadership transformationnel occupe dans la recherche actuelle. Pourtant, connaissant les liens étroits et parfois complexes que le leadership transformationnel et transactionnel entretiennent avec les autres variables psychologiques comme la personnalité ou les aptitudes, il n'est pas étonnant que les outils les plus récents, se proposent de mettre en parallèle des dimensions de personnalité et de leadership.

4. Les grands « audits » des qualités, compétences et personnalité des leaders

A. Les caractéristiques des leaders incompetents

Considérant qu'il est plus facile mais également plus utile de détecter les leaders incompetents, Hogan et Hogan (1997) ont proposé un inventaire le *Hogan Development Survey* basé sur une étude de mise en parallèle de dimensions de personnalité et de l'efficacité des leaders. Ils ont ainsi défini deux « côtés » de la personnalité : le côté lumineux (*bright side*) qui décrit les personnes lorsqu'elles sont à leur meilleur, et qui correspond aux dimensions des *Big Five* (Hogan, Curphy, & Hogan, 1994) et le « côté sombre » (*dark side*) qui caractérise les personnes lors de leurs échecs ou de mauvaises périodes et dépasserait les dimensions des *Big Five* (Hogan & Hogan, 2001).

Les caractéristiques associées au côté sombre seraient elles aussi dans le domaine de la personnalité normale, c'est à dire qu'elles ne rempliraient pas les critères nécessaires pour un diagnostic de trouble clinique, mais seraient tout de même fortement indésirables. Ainsi, le *Hogan Development Survey* a été conçu pour permettre d'identifier la présence de traits de personnalité pré-morbides caractéristiques des troubles de personnalité de l'axe II, tels que défini par le DSM (*Diagnostic and Statistical Manual of Mental Disorders*).

Les auteurs proposent un outil composé de 11 dimensions, organisés en trois groupes :

- ✦ S'éloigner des autres (*Moving away from people*)
- ✦ Aller contre les autres (*Moving agaonst people*)
- ✦ Aller vers les autres (*Moving towards people*).

Chacune de ces dimensions du HDS peut être mise en relation directe avec les troubles de la personnalité présenté dans le DSM (Hogan & Hogan, 2001) tel que présenté dans le tableau U.1.

DSM-IV Trouble de la personnalité :	Groupes de la HDS	Dimensions de la HDS
Borderline	S'éloigner des autres	Susceptibilité émotionnelle
Évitante		Précautionneux
Schizoïde		Réservé
Paranoïaque		Sceptique
Passive-Agressive		Trop sûrs de soi / Dédaigneux des autres
Antisociale	Aller contre les autres	Séducteur / Calculateur
Histrionique		Théâtral
Narcissique		Arrogant
Schizotypique		Imaginatif / Fantaisiste
Obsessionnelle - Compulsive	Aller vers les autres	Diligent
Dépendante		Respectueux / Attentif

Tableau U.1 :Description des échelles du HDS en lien avec les troubles de personnalité du DSM.

Issue de : Hogan, R., & Hogan, J. (2001). Assessing leadership: A view from the dark side. *International Journal of Selection and Assessment*, 9, 40.

Une grande étude sur les liens entre ces dimensions et différentes descriptions comportementales des managers, a permis de montrer les liens directs entre certaines de ces caractéristiques personnelles et des lacunes managériales.

Enfin, ce questionnaire ayant été élaboré au regard du DSM et avec l'objectif de détecter des traits reliés aux troubles de personnalité, il est peu étonnant de trouver des liens consistants entre des dimensions du HDS et des traits de personnalité.

Cependant, étant uniquement orienté vers le diagnostic de dysfonctionnement, ce questionnaire en perd un de ses objectifs – positionner les leaders sur un continuum d'efficacité, voire de potentiel de développement.

B. Les grandes évaluations multi-trait multi-méthode

Comme nous l'avons vu le leadership est loin d'être un concept unidimensionnel ou indépendant. De plus en plus, les professionnels et les chercheurs proposent des outils multi-facettes, s'intéressant, d'une part, à une grande variété de traits (personnalité, intelligence émotionnelle, capacités cognitives, leadership...) et, d'autre part, en faisant intervenir des observateurs en plus du manager évalué. Nous aborderons à présent ces grands audits de leadership – leur utilisation sous forme d'évaluation à 360°, sera quant à elle, abordée dans le chapitre suivant.

1. *Profilor*

Le Profilor (Hezlett, Ronnkvist, Holt, & Hazucha, 1997) est un outil conçu pour être utilisé dans le cadre des évaluations à 360° et vise à évaluer les compétence comportementales généralement nécessaires aux managers.

Le Profilor a été conçu sur la base d'une large revue de la littérature psychologique et du management et d'une analyse exhaustive du MPS (*Management Skills Profile* - Sevy, Olson, McGuire, Frazier, & Paaganen, 1985) grand outil des pratiques managériales composé de 18 dimensions regroupées au sein d'un modèle des compétences managériales en 8 parties (Administration, Leadership, Communication, Adaptation personnelle, Motivation et engagement, Connaissances techniques et aptitude cognitive)

Le Profilor, à la différence du MPS est composé de 130 comportements managériaux regroupés en 24 échelles (dont celles du MPS), organisées en quatre groupes :

- ⤴ Leadership centré sur les idées (*Thought leadership*)
- ⤴ Leadership centré sur les résultats (*Results leadership*)
- ⤴ Leadership centré sur les gens (*People leadership*)
- ⤴ Leadership centré sur soi (*Self leadership*).

Le **leadership centré sur les idées** se compose de comportements d'analyse efficace de la situation, de recherche active du changement, de la mise en place de plans concrets et de la connaissance de son domaine d'activité. Il requiert également la gestion de la production, l'utilisation de son expertise technique, une capacité à donner une direction à

l'action des autres et enfin l'utilisation de son bon sens.

Le **leadership centré sur les résultats** repose sur des comportements de recherche active de résultats, de leadership courageux face aux difficultés et sur la capacité du leader à partager son investissement dans le travail.

Le **leadership centré sur les gens** est quant à lui, constitué de comportements centrés sur la relation, de coaching, de compréhension mutuelle, de promotion de la communication ouverte et du travail d'équipe. Ce leadership est également composé par des comportements d'influence, de motivation et d'écoute des autres. Enfin, il requiert une capacité à gérer les conflits et à s'exprimer clairement.

Enfin le **leadership centré sur soi** comprend des actions d'intégrité et d'adaptabilité du leader, ainsi que des comportements faisant preuve de sa capacité à se développer et à apprendre.

Profilor possède une bonne consistance interne (de .82 à .84) et a été objet d'études de validation aussi bien sur le continent américain qu'en Europe (Robie, Johnson, Nilsen, & Hazucha, 2001).

2. Benchmarks

Benchmarks (Lombardo & McCauley, 1994) est un outil développé par le *Center for Creative Leadership* en Caroline du Nord. Cet outil a été construit sur la base d'interviews et d'enquêtes au cours desquelles des managers ont décrit les expériences clefs de leur carrières et les leçons qu'ils en ont tirés.

L'outil est composé de 106 items regroupés en 16 dimensions elles-mêmes organisées en trois groupes :

- ♣ Gestion de l'organisation (*Leading the organisation*)
- ♣ Gestion des autres (*Leading others*)
- ♣ Gestion de soi (*Leading yourself*)

Le groupe **gestion de l'organisation** est composé de quatre dimensions :

- ⤴ Perspective stratégique
- ⤴ Prise de décision
- ⤴ Gestion du changement
- ⤴ Capacité à rapidement saisir les choses (*being a quick study*).

Le groupe **gestion des autres** comporte, sept dimensions :

- ⤴ Mener les employés
- ⤴ Faire face aux problèmes des employés
- ⤴ Management participatif
- ⤴ Collaboration
- ⤴ Compassion et sensibilité
- ⤴ Facilité pour mettre les autres à l'aise
- ⤴ Respect des différences

Enfin, le groupe **gestion de soi** est composé, lui, de cinq dimensions :

- ⤴ Prise d'initiative
- ⤴ Maîtrise de soi
- ⤴ Équilibre entre la vie personnelle et professionnelle
- ⤴ Conscience de soi
- ⤴ Gestion de carrière.

Un quatrième groupe de facteurs est disponible pour le test *Benchmarks* et permet de mettre à jour des problèmes pouvant bloquer une carrière. Ce groupe vise à éclaircir les zones potentiellement problématiques chez le manager, il est composé de cinq dimensions :

- ⤴ Problèmes avec les relations inter-personnelles
- ⤴ Difficultés à construire et mener une équipe
- ⤴ Difficultés d'adaptation au changement
- ⤴ Incapacité à atteindre les objectifs
- ⤴ Orientation fonctionnelle trop limitée

Utilisé pour les évaluations à 360°, le *Benchmarks* est complété par le manager lui-même mais aussi par ses observateurs. L'ensemble des évaluateurs doivent noter dans quelle mesure le manager évalué présente les caractéristiques décrites dans chaque item. Les

réponses se font sur une échelle de type Lickert en cinq points, allant de « pas du tout » à « beaucoup ». Le questionnaire présente de très bonnes qualités métriques (Lindsey, Homes, & McCall, 1987), avec une consistance interne allant de .75 à .97.

La version française de *Benchmarks* à été réalisée par Claude Lévy-Leboyer. Cet outil, appelé en français « Repères de Carrière », a été le premier outil d'évaluation du leadership à 360°, en français.

3. Le Tryptique

Crée par Claude Lévy-Leboyer (2001) le Triptyque est un très bon exemple d'outil multi-trait et un des rares développé entièrement en français.

Le test évalue les compétences managériales, des dimensions de la personnalité au travail et des aptitudes cognitives en trois modules indépendants. Les comportements doivent être évalués selon la fréquence de leur observation sur une échelle en cinq points.

Sept dimensions sont évaluées dans le module de **compétences managériales** :

- ✦ Prise de décision
- ✦ Planification stratégique
- ✦ Gestion de l'action
- ✦ Communication
- ✦ Gestion de l'équipe
- ✦ Relations humaines
- ✦ Concevoir et communiquer une vision

Six dimensions sont évaluées dans la partie **personnalité au travail** :

- ✦ Ambition
- ✦ Stabilité émotionnelle
- ✦ Besoins sociaux
- ✦ Ouverture intellectuelle
- ✦ Fiabilité
- ✦ Convivialité

Enfin sept autres dimensions font partie du groupe **aptitudes cognitives** :

- ⤴ Pensée critique
- ⤴ Structuration d'informations
- ⤴ Créativité
- ⤴ Expression écrite
- ⤴ Expression orale
- ⤴ Compréhension écrite et orale
- ⤴ Aptitude numérique.

La version Triptyque est destinée aux dirigeants ou cadres confirmés. Une version de l'outil pour cadres du *middle management* est également disponible sous le nom Compétences. Le Triptyque tout comme Compétences, permettent une évaluation en 360° : ils proposent donc par défaut deux versions du questionnaire – une pour le manager et une pour ses observateurs. Publié par les ECPA (Editions du Centre de Psychologie Appliquée) cet outil est peu présent dans les publications ou les recherches.

Cependant, malgré leur richesse, aucun des questionnaires étudiés précédemment ne permet d'approcher le niveau d'analyse inhérent à l'approche clinique du leadership. En effet, l'approche clinique du leadership recherche avant tout à mettre en lumière les dimensions sous-jacentes et profondes du comportement des leaders, le GELI, se propose d'approcher ce niveau d'analyse.

4. Le GELI (*Global Executive Leadership Inventory*)

Le GlobelInvent (*Global Leadership Life Inventory*), devenu par la suite le *Global Executive Leadership Inventory* (Kets de Vries et al., 2004) est un instrument d'évaluation du leadership conçu dans le cadre d'un large projet s'intéressant aux actions et aux comportements quotidiens des managers efficaces. Cet outil vise à pallier un manque certain de la gamme pourtant large des outils d'évaluation du leadership – le manque de profondeur des dimensions mesurées.

En effet, de sensibilité psychanalytique, le Pr. Manfred Kets de Vries visait tout d'abord à aider les managers de haut niveau à déterminer les zones de leadership qui nécessiteraient un travail de développement personnel. Or, la plupart des outils existants ne proposant qu'une évaluation superficielles des qualités de leadership, il a décidé de mettre au point le GELI, qui lui, opte pour une approche plus clinique du leadership et

permet de mettre en lumière le « théâtre intérieur » des leaders.

Plusieurs étapes ont été suivies afin d'isoler les dimensions et construire une première version du test. Tout d'abord, des entretiens semi-structurés ont été menés sur une période de trois ans avec des managers expérimentés, en abordant à travers des questions ouvertes la notion de leadership « exemplaire ». Sur la base de ces résultats, et des recherches antérieures, des thèmes généraux ont pu être isolés puis testés dans le cadre des études de pré-test du questionnaire pilote.

Trois études ont été réalisées pour éprouver le questionnaire une, auprès de Directeurs Généraux en activité suivant un programme de formation à l'INSEAD, la seconde sur un échantillon d'étudiants en MBA et la troisième auprès de managers seniors inscrits à un programme en gestion.

Les analyses de ces données ont permis de confirmer l'existence de 12 facteurs robustes avec une forte consistance interne.

Utilisé lors des programmes de développement managérial, notamment à l'INSEAD (*The Challenge of Leadership: Creating Reflective Leader*) mais également au sein des entreprises, le GELI se positionne en tant qu'élément diagnostique initial.

Ayant fait le constat, comme beaucoup d'autre, que la perception que les leaders ont d'eux-mêmes est souvent différente de celle que leur entourage personnel et professionnel a d'eux, le GELI propose, comme les outils abordés précédemment, une version observateur, permettant la mise en place d'une évaluation en 360°.

La version finale du GELI est composée de 12 dimensions, les exemples d'items présentés ci-après, en français, sont issus de la version française du GELI et correspondent aux items anglais de l'article de validation (Kets de Vries et al., 2004). Les questionnaires self et observateur sont présentés dans l'annexe 1.

▲ ***Visioning*** – Donner une vision :

Les dirigeants doivent être capables de percevoir et d'agir sur les principales tendances de l'environnement, car ils se doivent d'être orientés vers l'avenir. Ils dirigent et guident, structurant les informations complexes de façon à les rendre accessibles à leurs subordonnés. Ils doivent être prêts à remettre en question le *statu quo* et être ouverts au changement.

Un manager présentant un score élevé sur cette dimension, s'efforce de saisir les opportunités et est capable de remettre en cause le *statut quo*. De même, il parvient à avoir une vue d'ensemble, et simplifiée, de situations complexes.

Un exemple d'item pour cette dimension :

« Je donne envie à mes collaborateurs de voir au-delà des limites existantes. ⁵»

✧ **Empowering – Donner le pouvoir**

Les dirigeants doivent exprimer des attentes élevées et avoir confiance dans leurs employés. Connaissant les dangers de la rétention d'information, ils doivent informer leur personnel régulièrement, les motiver. Dans les prises de décision, ils doivent s'assurer que chaque personne ayant un intérêt dans une décision, soit partie prenante dans le processus de prise de décision, afin de créer un sentiment d'appartenance. Également, ils doivent exprimer clairement ce qu'ils attendent de ceux qui les suivent, tout en considérant les erreurs comme des opportunités pour apprendre.

Un manager présentant un score élevé sur cette dimension informe et minimise la rétention d'information, délègue, s'efforce de créer un sentiment d'appartenance et tolère les erreurs.

Un exemple d'item pour cette dimension :

« J'essaie d'impliquer mes salariés dans la prise de décisions. ⁶»

✧ **Energizing – Donner de l'énergie**

Les dirigeants doivent canaliser l'énergie de ceux qui les suivent d'une façon constructive pour atteindre les buts fixés. Ils doivent être proactifs, orientés vers l'action et passionnés par ce qu'ils font. Diriger par l'exemple : ils font ce qu'ils disent. Par leur enthousiasme, ils parviennent à mobiliser leurs employés pour que les choses soient faites.

Un manager présentant un score élevé sur cette dimension parvient à la fois à mobiliser les gens et à les diriger par l'exemple, en étant un modèle.

⁵« I inspire my people to look beyond existing boundaries. »

⁶« I always try to involve my employees in decision-making. »

Un exemple d'item pour cette dimension :

« Je mobilise mes collaborateurs pour que les choses soient faites. ⁷»

▲ **Designing and aligning – Conception et contrôle**

La conception et le contrôle composent le rôle "architectural" des managers exemplaires. Les dirigeants doivent à tout prix donner à leur vision un fondement structurel. Au quotidien, ils se doivent de tenir les autres responsables de leurs engagements et délais, tout en étant orienté vers le résultat. Enfin, en tant que garant de la culture d'entreprise, ils veillent aussi à ce que le comportement de leurs salariés soit en accord avec les valeurs culturelles.

Un score élevé sur cette dimension décrit un manager capable de faire correspondre les valeurs, les attitudes et les comportements de ses salariés avec les systèmes mis en place. C'est également un manager tenant ses employés responsable des engagements et des délais pris par eux.

Un exemple d'item pour cette dimension :

« Je définis pour mes collaborateurs des normes et objectifs de performances précis. ⁸»

▲ **Rewarding and feedback – Récompense et feedback**

Les dirigeants doivent savoir récompenser leur personnel. Quel que soit le système de récompenses mis en place pour motiver les salariés – stock options, primes, avantages, plans de participation aux résultats... – ils doivent s'assurer que celui-ci est juste. A leur niveau, ils veillent également à féliciter les salariés pour leurs réussites, grandes et petites, et à récompenser de façon exceptionnelle, les performances remarquables. Aussi, les managers s'efforcent de guider leurs salariés en donnant un feedback constructif.

Un score élevé sur cette dimension décrit un manager qui s'assure que les encouragements, récompenses et sanctions soient justes tout en donnant régulièrement un feedback constructif et efficace.

Un exemple d'item pour cette dimension :

« Je veille à ce que les réussites des collaborateurs soient reconnues. ⁹»

⁷« I mobilize people to get things done. »

⁸« I set clear performance standards and goals. »

⁹« I make sure that achievements are recognized along the way. »

✧ **Teambuilding – Création d'équipe**

Les dirigeants doivent mettre en place des partenariats et créer un sens de l'engagement parmi leurs salariés de sorte que le groupe, dans son ensemble, puisse réaliser avec efficacité ses objectifs. Les managers s'assurent d'impliquer activement les autres dans le processus de prise de décisions, gèrent les conflits et recherchent des solutions gagnant/gagnant lorsque des dissensions apparaissent. Ils doivent favoriser la coopération et mettre en place une atmosphère de confiance en plaçant l'intérêt du groupe avant leurs buts personnels.

Un score élevé sur cette dimension décrit un manager qui encourage une résolution de conflits constructive, qui crée une atmosphère coopérative tout en faisant preuve d'un comportement loyal envers l'entreprise. Enfin, ce manager considère la diversité des membres du groupe comme un avantage.

Un exemple d'item pour cette dimension :

« J'essaie de veiller à ce que tous les membres du groupe sentent qu'ils participent au processus décisionnel. ¹⁰»

✧ **Outside orientation – Orientation vers l'extérieur**

Pour la performance de l'entreprise, des relations de confiance et de respect mutuel avec les actionnaires et les partenaires est primordiale. Les managers sont donc très attentifs à répondre aux exigences de leurs clients et ils restent très proches de tous leurs partenaires. Ils sont également conscients de leur responsabilité sociale envers la communauté à laquelle ils appartiennent.

Un score élevé sur cette dimension décrit un manager qui participe à la bonne gestion des relations avec les clients et les partenaires extérieurs.

Un exemple d'item pour cette dimension :

« Je veille à ce que nous concentrions tous nos efforts sur la satisfaction du client. ¹¹»

¹⁰« I make sure that all participants feel that they contribute to to the decision-making process. »

¹¹« I make sure that customer satisfaction stands central. »

✧ **Global mindset – État d'esprit « Global »**

Les dirigeants doivent avoir une excellente connaissance de la situation socioéconomique et politique dans leur ensemble. Ils doivent être curieux de ce qu'ils ne connaissent pas mais également intéressés de connaître les autres cultures. A l'aise dans les situations culturelles compliquées, ils doivent être capable de gérer et de faire partie d'une équipe multiculturelle.

Un score élevé sur cette dimension décrit un manager qui a un état d'esprit « global », et qui s'intéresse aux autres cultures.

Un exemple d'item pour cette dimension :

« Je m'adapte avec facilité aux pratiques des affaires dans des cultures autres que la mienne. ¹²»

✧ **Tenacity – Ténacité**

Les dirigeants ont le courage et la ténacité de défendre des décisions impopulaires lorsqu'ils pensent qu'elles sont nécessaires. Ils doivent être prêts à prendre des risques pour atteindre l'excellence. Très résilients, ils ne doivent pas se décourager facilement.

Un score élevé sur cette dimension décrit une personne qui prend fortement appui sur ses convictions, qui est résiliente et qui n'abandonne pas facilement.

Un exemple d'item pour cette dimension :

« Je suis prêt à soutenir une décision impopulaire si je pense que c'est la bonne. ¹³»

✧ **Emotional intelligence – Intelligence Émotionnelle**

Les dirigeants doivent connaître l'importance d'une direction « en douceur » et s'efforcer d'élargir leur connaissance d'eux-mêmes, en s'engageant dans un processus permanent de réflexion sur eux. Bons gestionnaires de leurs émotions, les managers doivent connaître leurs employés et savoir composer avec les émotions des autres. Conscients de l'impact de leur propre comportement sur les personnes qui travaillent avec eux, ils utilisent activement le feedback pour s'améliorer et aider les autres à faire de même.

12« I am good at adapting to business practices in cultures other than my own. »

13« I am prepared to stick to an unpopular decision if I feel that it is the right one. »

Un manager avec un score élevé sur cette dimension est engagé dans un processus permanent de réflexion sur lui-même, tout en étant, à la fois, maître de lui. Il s'agit de quelqu'un capable de tirer des enseignements des erreurs et d'utiliser le feedback pour s'améliorer.

Un exemple d'item pour cette dimension :

« J'essaie de créer la confiance entre les collaborateurs avec lesquels je travaille. ¹⁴»

✧ **Life balance – Équilibre de vie**

Les dirigeants doivent avoir un style de vie équilibré. Il peut être risqué de « mettre tous leurs œufs dans le même panier », aussi ils doivent diversifier leurs centres d'intérêt. Ils savent définir leurs priorités, mais sont également conscients qu'il est important de se régénérer c'est pourquoi ils s'engagent activement dans des activités sans rapport avec le travail. Ils ont aussi dans leur entourage des confidents avec qui ils peuvent parler de choses personnelles.

Un manager avec un score élevé sur cette dimension donne de l'importance à son équilibre de vie, parvient à diversifier ses intérêts et a des personnes proches à qui se confier.

Un exemple d'item pour cette dimension :

« Je définis des priorités tant dans ma vie privée que professionnelle. ¹⁵»

✧ **Resilience to stress – Résilience au stress**

Les dirigeants font des efforts pour gérer le stress de leurs vies. Ils connaissent les différentes dimensions du stress et ils veillent à limiter leur niveau de stress au travail. Ils réalisent l'impact de leurs choix de carrière sur leur santé mentale et physique et comprennent, comment les pressions de la vie (problèmes familiaux, financiers..) affectent leur état mental.

Un manager avec un score élevé sur cette dimension parvient à gérer le stress dû à son travail, à ses choix de carrière et aussi celui dû à sa santé.

14« *I work to generate trust among my people.* »

15« *I set priorities in both my private and my professional life.* »

Un exemple d'item pour cette dimension :

« La pression au travail est devenue beaucoup trop stressante. ¹⁶»

Les items sont présentés sous forme d'affirmations, à évaluer sur une échelle de type Likert, en sept points, allant de « Jamais » à « Toujours ». Comme nous le verrons dans la partie Méthode (Chapitre 5), ce test présente d'excellentes qualités psychométriques.

¹⁶« *I feel a lot of pressure at work.* »

Nous avons détaillé dans ce chapitre le très large spectre des outils d'évaluation du leadership. Répondant souvent aux théories qui les fondent, bien des instruments ont été développés par les auteurs mêmes des modèles sous-jacentes. Ainsi, le *Managerial Grid Questionnaire* – permet de situer le manager sur la grille managériale de Blacke et Mouton alors que le *Leadership Behavior Description Questionnaire* de l'Université de l'Ohio, permet de positionner un leader en fonction de sa préférence pour les comportements d'initiation de structure ou de considération. De même, le *C-R Scale of Charismatique Leadership* propose de catégoriser le leader selon la théorie de l'attribution de Conger et Kanungo, alors que le *Multifactor Leadership Questionnaire* de Bass positionne un manager sur chacune des dimensions présentes dans le modèle du leadership transactionnel, transformationnel et laissez-faire.

Ensuite, nous éloignant un peu des approches directement basées sur une théorie spécifique, nous avons étudié les grands audits de leadership. Associant des dimensions de personnalité, d'intelligence émotionnelle ou d'aptitude cognitive, ces outils se proposent de décrire un leader sous tous les angle psychologiques. Pourtant seul le GELI permet d'approcher exhaustivement la complexité et les dimensions latentes à l'œuvre dans la psyché du leader.

Nous pouvons noter cependant une certaine redondance des outils. Ainsi, bien que les dimensions soient parfois nommées différemment, souvent des caractéristiques très similaires se retrouvent d'un questionnaire à l'autre. Certes, les arguments à l'origine de la création de nouveaux outils se justifient souvent au niveau théorique, pourtant dans la pratique, la multiplication des instruments de mesure ne permet pas au chercheur ou au professionnel de s'y retrouver facilement.

De même, qu'ils fassent partie des questionnaires de leadership les plus utilisés ou des outils moins populaires, la quantité d'études de validation rapportant des résultats contradictoires laisse perplexe.

Nous pouvons donc nous poser la question de la popularité de certains de ces questionnaires. Est-elle réellement justifiée par les qualités de ceux-ci ? Par leur approche théorique ? Ou bien est-ce lié à des raisons plus économiques. Par ailleurs, nous tenons également à attirer l'attention du lecteur sur l'utilisation de certains questionnaires dans les recherches. Le fait qu'un questionnaire donné soit souvent utilisé n'afflue pas nécessairement sur ses qualités psychométriques. Il importe que ses qualités métriques soient bien établies en amont, lors d'études de validation à

proprement parler.

Enfin, tout au long de ce chapitre, nous avons noté le nombre important de questionnaires proposant une évaluation en 360°. Ces questionnaires offrent au leader l'occasion de confronter sa vision de lui-même avec la vision que d'autres ont de lui.

Cette approche originaire du management anglo-saxon, a connu un succès important dans les autres pays européens dès les années 90. Or, la démarche d'évaluation à 360°, du fait de ses spécificités, nécessite plus encore que d'autres méthodes, des bases psychométriques solides. En effet, vu le type de comparaison et la manière dont les résultats des évaluations à 360° sont analysés, les questionnaires se doivent de présenter des qualités métriques exemplaires – qu'il s'agisse de consistance ou de structure interne. C'est cette approche à 360°, son origine, ses intérêts et ses limites que nous allons aborder maintenant.

Chapitre 4. Une approche en plein essor : les 360 degrés

Jusqu'à présente nous avons passé en revue les théories et les définitions du leadership, ses antécédents psychologiques mais aussi ses effets et la mesure de son efficacité. Puis nous nous sommes intéressés à la mesure du leadership à proprement parler. Qu'il s'agisse d'outils se proposant de mesurer les styles de leadership ou les traits particuliers d'un modèle donné, de plus en plus d'instruments de mesure posent la question de l'hétéro-évaluation. En effet, être un leader, c'est aussi être perçu et jugé comme tel.

Nous allons donc dans ce quatrième chapitre nous intéresser à l'hétéro-évaluation. Cette méthode, popularisée dans le cadre des évaluations à 360°, a connu une utilisation croissante depuis les années 90. Nous aborderons dans la première partie de ce chapitre l'origine de cette approche, ses caractéristiques et son intérêt dans l'évaluation du leadership et des qualités managériales au sens large. Puis, dans la seconde partie, nous mettrons en lumière les questions méthodologiques posées par cette approche, les traitements méthodologiques qu'elle requiert et les limites des approches psychométriques classiques du fait de sa structure même.

1. De l'auto-évaluation à l'évaluation par les autres

A. L'auto-évaluation

Au début de la psychométrie, les épreuves destinées à mesurer les aptitudes intellectuelles des sujets se basaient sur la réalisation d'un certain nombre de tâches (situations/problèmes) par les sujets. Leur réussite ou leur échec, le niveau de difficulté de la tâche et le niveau de réussite moyen estimé au niveau de l'échantillon permettait de situer ces individus dans la population générale en fonction de leur performance. Celle-ci était donc qualifiée de faible, moyenne ou élevée.

Par la suite, une démarche similaire à été mise en place pour évaluer la personnalité. En effet, les questionnaires de personnalité visent à positionner un individu sur un trait latent (comme l'extraversion ou l'agréabilité) à partir de ses réponses à une série d'items représentatifs des comportements associés à ce trait.

Or c'est là qu'apparaît l'une des grande différence entre la mesure des aptitudes et celle des dimensions conatives et de personnalité : dans les mesures de personnalité, l'individu exprime sa perception de son comportement. Il ne s'agit donc plus de juger la réussite ou l'échec réel d'un individu à un problème donné, mais la perception de l'individu sur ses capacités. C'est de cette différence que viennent les nombreuses difficultés (biais) suscitées par ces auto-évaluations (Huteau et Lautrey, 1999).

En effet, plusieurs biais apparaissent dès lors que l'on parle d'auto-évaluation. L'individu est-il capable d'avoir une vision claire et précise de lui-même ? N'aura-t-il pas tendance à sur-évaluer ses capacités pour répondre aux attentes de la société (désirabilité sociale) ? Ou bien même, aura-t-il l'envie d'être honnête ?

Ces questionnements ont donné lieu à un arsenal de méthodes statistiques et méthodologiques visant à limiter les biais présents dans l'auto-évaluation. Cependant, une autre catégorie d'approches à choisi de ne pas se limiter à l'auto-évaluation.

B. L'Hétéro-évaluation

Comme nous l'avons dit lors de la présentation des outils de mesure du leadership, nombre d'entre eux choisissent d'évaluer les leaders non seulement à travers leur propre perception d'eux-mêmes mais aussi à travers la perception qu'en ont leurs subordonnés et/ou les personnes qui travaillent avec eux.

Bien que cette approche puisse paraître nouvelle en psychométrie et en psychologie du travail, d'autres branches de la psychologie l'utilisent également. Ainsi, en psychologie du développement ou en psychopathologie, pour des problématiques aussi différentes que : la perception des symptômes (Barkley, Knouse, & Murphy, 2011), la présence de pathologies chez l'enfant (De Los Reyes & Kazdin, 2005), la perception des conduites asociales (Gresham, Elliott, Cook, Vance, & Kettler, 2010), l'éducation des enfants (Schwarz, Barton-Henry, & Pruzinsky, 1985; Tein, Roosa, & Michaels, 1994) ou la description des difficultés d'apprentissage (Cole, Maxwell, & Martin, 1997) l'hétéro-évaluation est très populaire.

Travaillant sur les problématiques de la mesure liées à l'observation, dans les années 80, certains auteurs ont suggéré qu'agrégé des scores issus de rapporteurs multiples (différents observateurs) sur un même objet (un même trait d'un même individu) pouvait

permettre de passer outre les biais des auto-évaluations et ainsi capturer une plus grande partie de la variance vraie et donc améliorer la généralisabilité des mesures (Rushton, Brainerd, & Pressley, 1983).

Nous verrons dans la partie d'analyse dans quelle mesure ce constat est également valable pour les évaluations à 360° mais aussi les questions que cette démarche pose.

Un peu plus tard, dans les années 90, d'autres auteurs font remarquer qu'une simple agrégation des scores pourrait, en fait, agréger des erreurs systématiques – ils suggèrent donc d'utiliser des modèles d'équation structurels pour évaluer la fidélité et les biais présents dans les évaluations (Cook & Goldstein, 1993; Tein et al., 1994).

Ainsi, dans les études portant sur les relations intra-familiales, utilisant l'approche multi-trait, multi-méthode, Jacob et ses collaborateurs (1987) ont réalisé une méta-analyse sur les études empiriques. Ils ont repéré une relations très faibles entre les différentes méthodes (c'est à dire entre les mesures des observateurs, des selses et des conjoints). Bien qu'ils aient essentiellement conclu à la nécessité d'une plus grande rigueur des mesures d'autres à leur suite, n'ont plus associé ces faibles liens avec de simples biais dans les mesures. Ainsi, Furman et Jacob (1997) ont mis en avant la possibilité de différences réelles existant dans les perspectives de chacun des évaluateurs. Il devient donc possible d'envisager que les différentes mesures correspondent à des évaluations de construits quelque peu différents, même s'ils sont rattachés à une même dimension sous-jacente.

De même, dans le domaine de la de la psychopathologie, les auteurs tendent à trouver des liens témoignant d'accords modestes (Achenbach, 1991; Biederman et al., 1993) entre les différentes sources d'évaluation des troubles d'un enfant donné. De fait, les professionnels accordent différentes valeurs aux informations fournies par chacune des sources selon la pathologie évaluée. Ainsi, des recherches ont montré que l'auto-évaluation par le jeune est la source la moins prédictive dans les cas d'hyperactivité ou d'inattention ; alors que celle des enseignants est la moins utile en cas de problèmes plus internalisés (Loeber, Green, Lahey, & Stouthamer-Loeber, 1991).

Une des grandes difficultés de la psychopathologie reste de déterminer la part des différences dues aux outils et celle due aux contextes, extrêmement variés, dans lesquels les divers observateurs interagissent avec le sujet. Ainsi, des parents, des enseignants et des médecins ne côtoient clairement pas le sujet dans les mêmes environnements et, qui

plus est, ils utilisent souvent des outils différents pour évaluer des comportements, eux aussi, différents.

Comme nous le verrons dans la partie qui suit, les méthodes d'évaluation à 360° utilisées pour l'évaluation du leadership ne portent, elles, que sur l'environnement professionnel. De même, bien que certaines recherches utilisent des observateurs très divers, l'approche la plus classique se limite à des observateurs interagissant avec le sujet sur son lieu de travail. Enfin, en ce qui concerne ces outils, le 360° utilise systématiquement les mêmes outils de mesure pour tous les observateurs.

2. Le 360 : une présentation

A. La définition

On considère une évaluation à 360°, dans les situations où une évaluation est réalisée d'une part, par le sujet concernée (les *selves*), et, d'autre part par un ensemble de personnes proches de lui / elle (dénommés de manière classique les observateurs). Il est d'usage d'appeler *self* (au pluriel *selves*) le questionnaire renseigné par le leader lui-même et correspondant à son auto-évaluation. Les répondants aux hétéro-évaluations sont qualifiés d'observateurs (anglais : *observers*). Dans le 360°, une évaluation est donc constituée d'un ensemble de protocoles recueillis auprès d'un leader comprenant son auto-évaluation (*self*) et les hétéro-évaluations de ses observateurs.

Toutes ces évaluations ont pour objectif de mettre à jour les écarts et les disparités existant entre la perception que le sujet a de lui-même et celles que ses observateurs ont de lui.

D'une manière générale on peut distinguer plusieurs catégories d'observateurs tels que cités par Lévy-Leboyer (2000) :

- ♣ « le supérieur hiérarchique » la personne à qui on répond ou à qui on rend compte.
- ♣ « les subordonnés » ils sont susceptibles d'apporter des informations précieuses parce qu'ils ont souvent des occasions plus diversifiées d'observer leur patron.
- ♣ « les collègues »
- ♣ « les clients » internes ou externes.
- ♣ « les fournisseurs » ils observent le leader dans une situation de contrainte.
- ♣ « les amis, membres de la famille » ils ont l'occasion de voir le participant sous des angles et dans des situations différentes de celles du travail.

L'un des objectifs du 360 – comme nous le verrons par la suite – est donc bien de croiser des perceptions d'observateurs interagissant avec le *self* dans des situations professionnelles différentes.

Sphères d'observation	Manager	Pairs	Subordonnés	selves
<i>Performance à la tâche</i>				
Comportement	Occasionnel	Fréquent	Rarement	Toujours
Résultats	Fréquent	Fréquent	Occasionnel	Fréquent
<i>Aptitudes inter-personnelles</i>				
Comportements	Occasionnel	Fréquent	Fréquent	Toujours
Résultats	Occasionnel	Fréquent	Fréquent	Fréquent

Tableau V.1 Fréquence des observations par sphères et par type d'observateur.

Tiré de Borman, W. C. (1974) The rating of individuals in organisations : an alternative approach. *Organisational Behaviour and Human Performance*, 12, 105-124.

Mises en place dans un premier temps pour pallier les limites des évaluations *Top-down* classiques des managers (évaluations par les supérieurs de leurs subordonnés), les approches à 360° ont fait leur apparition lors des évaluations annuelles de la performance (*Performance appraisals*) réalisées par les employeurs.

B. Les avantages et les limites

L'utilisation massive et le grand attrait des méthodes d'évaluation par l'approche à 360° peuvent s'expliquer par un ensemble d'avantages. Elles sont perçues comme efficaces car les informations qu'elles fournissent proviennent de sources multiples et de personnes dont l'opinion est importante (Hazucha, Hezlett, & Shneider, 1993). Cette information acquiert donc une plus grande crédibilité et par conséquent augmente la motivation de l'individu évalué pour changer son comportement (Hazucha, Latham, & Wexley, 1982; Hellevik, Hazucha, & Shneider, 1992).

Ainsi, d'après De Nisi et Griffin (2001), les principaux avantages des évaluations à 360° sont, d'une part, de donner des informations sur les perceptions de soi par les autres, permettant par là-même au manager évalué de s'améliorer (en mettant l'accent sur ses faiblesses). D'autre part, que ces informations soient issues de différentes sources ayant chacune un rapport spécifique avec le sujet.

Dans le cadre d'une utilisation en entreprises, certains auteurs ajoutent d'autres avantages comme la mise en lumière des dimensions de performance importantes pour l'entreprise, le changement de la culture d'entreprise et l'attention portée au service client (Antonioni, 1996; London & Beatty, 1993; Morgeson, Mumford, & Campion, 2005).

1. Les sources d'évaluation

Chaque source d'évaluation présente des biais et des limites qui lui sont propres en voici une description adaptée du livre *Psychological Assessment in the Workplace* de Cook et Cripps (2005) :

- ✧ **Le supérieur hiérarchique** est la personne qui traditionnellement fait les évaluations or, en pratique, le supérieur voit peu du travail effectif de son subordonné car en plus du manque d'occasions il a également un grand nombre de tâches et missions qui lui sont propres.

Qui plus est, il n'est pas exempt de biais. En effet, des éléments idiosyncrasiques (ne pas aimer qu'on porte des chaussures rouges, qu'on ne se tienne pas droit...) mais aussi des stéréotypes et des biais de groupe (contre les femmes, les personnes âgées...) tendent à diminuer l'exactitude de l'évaluation.

- ✧ **Le N+2**, sollicité dans un certain nombre d'évaluations à 360°, il à l'avantage de sa vision globale (stratégie générale de l'entreprise, objectifs d'ensemble...) et peut percevoir la contribution de chacun.

En même temps, n'ayant que peu d'occasions d'interagir directement avec le manager évalué, il risque de baser son évaluation sur la réputation, les oui-dires et les échos dont il a connaissance.

- ✧ **Les collègues et les pairs**, sont des personnes effectuant les mêmes tâches que l'évalué et ayant donc l'opportunité de l'observer dans son quotidien. Aussi, ayant souvent un grand nombre de collègues (et un seul manager) le poids accordé à l'avis de ce groupe prend souvent plus de valeur en vue du changement (plus de précisions seront apportées dans la seconde partie concernant l'accord du groupe).

Pourtant, outre la difficulté pouvant provenir d'une mauvaise information entourant l'évaluation (sensation de dénoncer, de « casser du sucre sur le dos »...) un phénomène

de lynchage peut également s'observer (London & Beatty, 1993; Morgeson et al., 2005). En effet, l'ensemble des évaluations, quand elles sont négatives, sont rarement indépendantes (tous n'aiment pas la personne évaluée, certains collègues meneurs n'aiment pas la personne évaluée, d'autres sont jaloux de sa réussite...) on retrouve donc souvent un effet de Halo sur l'ensemble des évaluations.

- ✧ **L'évaluation par les subordonnés** est un grand atout de la méthode à 360° dans la mesure où ce groupe particulier, donne un éclairage nouveau sur la performance ou la réussite d'un individu.

Pourtant l'évaluation « *bottom-up* » partage certains inconvénients avec celle faite par les pairs. En effet, un phénomène de lynchage n'est pas à exclure, il est d'ailleurs encore plus facile à imaginer pour les subordonnés.

- ✧ **L'auto-évaluation** n'a pas toujours été partie prenante des approches à 360°. Elle a l'avantage d'impliquer l'évalué, de réduire son système de défense et enfin d'accroître sa compréhension de ses points faibles et de lui-même.

Pourtant des biais sont tout de même nombreux et d'après Fletcher (1997), il importe de recourir à l'auto-évaluation dans le cas où les autres évaluateurs ont peu d'occasion d'observation, quand l'évaluation a uniquement un but développemental et lorsqu'il n'y a pas de comparaison directe avec d'autres évalués.

2. L'indulgence

De manière générale, lors d'une évaluation à 360°, on observe plus de « sur-performants » que de « sous-performants ». Il semble donc exister un effet général d'indulgence et plusieurs éléments tendent à l'accroître.

La restitution face à face tend à rendre l'évaluation de l'observateur plus indulgente. La préoccupation et l'importance accordée par l'évaluateur à une bonne atmosphère de travail, tend également à augmenter l'effet d'indulgence. En effet, un mauvais score risque de détériorer l'atmosphère de travail pour les jours voir les semaines suivant l'évaluation.

L'effet de préservation de soi est également un élément pouvant contribuer à l'indulgence de l'observateur. De même que l'atmosphère de travail, les mauvaises évaluations

pourraient entraîner de l'agressivité de la part de l'évalué, l'évaluateur choisirait donc d'être indulgent pour limiter les commentaires négatifs de son évaluation.

De même, lorsque les évaluations sont réalisées dans un but d'évaluation, les observateurs tendent à être plus indulgents que dans les 360° réalisés dans un but développemental.

Au niveau interculturel, plusieurs recherches (Atwater, Wang, Smither, & Fleenor, 2009; Farh & Cheng, 1997; Farh, Dobbins, & Cheng, 1991; Xie, Roy, & Chen, 2006; Yik, Bond, & Paulhus, 1998) ont également montré un effet d'indulgence plus marqué (pour les selves, les pairs et les subordonnés) dans les cultures dites Individualistes¹.

Enfin, l'effet des « départements à fort prestige » a également pu être observé dans les évaluations. Ainsi, dans l'*US Navy* (Murphy & Cleveland, 1995) l'inflation des évaluations s'accroît avec le prestige des unités d'appartenance. Les membres des unités nucléaires sont donc systématiquement mieux notés que les techniciens à terre par exemple. Les auteurs expliquent cet effet, par l'effet d'attente des performances : si un sujet est membre d'un corps d'élite, cela veut forcément dire qu'il est exceptionnellement performant.

3. L'effet de Halo

Quelles que soient les dimensions évaluées dans les questionnaires, la plupart des auteurs observent des « effets de Halo » sur les résultats.

Ainsi, les « bons employés » seraient bons dans tous les domaines, alors que les « mauvais », le seraient dans tous les domaines également. Cet effet n'est pas spécifique aux questionnaires à 360°, il s'observe également dans d'autres types d'évaluation comme les cotations d'entretien, les références... Pourtant l'effet de Halo est particulièrement gênant pour les 360° car leur objectif est développemental et vise à mettre en évidence des zones d'amélioration et de réussite. Or, du fait de cet effet, l'ensemble des dimensions tendent à présenter des scores similaires, cet objectif se trouve donc contrarié.

¹ En comparant les biais de modestie observée, des résultats plus marqués ont été trouvés pour les pays à culture collectiviste (Chine, Taiwan) comparativement aux pays à culture individualiste (Amérique du Nord et Europe).

Les 360 ne sont donc pas sans poser de problèmes. La très grande quantité d'informations qu'ils permettent d'obtenir rend en même temps difficile pour les managers évalués de la traiter efficacement.

Par ailleurs, au vu de leur richesse, l'interprétation des résultats nécessite de l'aide extérieure pour comprendre comment traiter les informations, parfois contradictoires. En effet, il est difficile de réconcilier les différences d'appréciation entre les évaluateurs qui sont non seulement en désaccord avec l'auto-évaluation du manager mais également en désaccord entre eux.

Enfin, la mise en lumière de ces zones de désaccord tend à créer un inconfort et parfois une incompréhension chez le manager évalué. Ceux-ci ressentent donc souvent le besoin d'un suivi et d'un accompagnement de la part de l'évaluateur pour ne pas « simplement » se laisser démoraliser par les résultats, mais mettre en marche le changement.

C. L'objectif du 360 : le changement

Créées pour être utilisées dans le cadre de sessions de développement personnel et de développement managérial, les évaluations à 360° visent à provoquer une prise de conscience en vue d'un changement chez le manager évalué.

Comme les informations (commentaires positifs et négatifs) des évaluations à 360° proviennent de sources différentes et nombreuses, la personne évaluée tend à accepter plus facilement les commentaires reçus. Qui plus est, des recherches antérieures ont montré que le simple fait de recevoir des informations sur ses performances (du *feedback*) permet d'améliorer ces dernières (Hegarty, 1974). Amélioration qui, selon les études, peut se maintenir jusqu'à deux ans après la réception du *feedback* (Hazucha et al., 1993).

De même, grâce à l'auto-évaluation, le manager devient conscient des différences entre sa propre évaluation et celles des observateurs. Pour Yammarino et Atwater (1997) cette étape est la première qui mènera au changement. Ainsi, grâce aux formations et à l'accompagnement dont il bénéficiera, le manager pourra tenter d'éliminer ces différences (Waldman, Atwater, & Antonini, 1998) et conformément à la théorie du contrôle, les différences entre sa propre perception de sa performance et celle des autres sera un facteur de motivation important pour engendrer une progression (Tornow & London, 1998).

1. Facteurs favorisant le changement

McCauley et Hughes-James (1994) ont réalisé une étude auprès de directeurs d'école. Dans leurs conclusions les auteurs montrent que plus de la moitié des directeurs ayant le plus modifié leurs comportements étaient ceux qui avaient rencontré le plus de difficultés lors de leur entrée en fonction. Ils précisent cependant que ce changement ne s'était avéré effectif que dans le cas d'un fort soutien familial.

De même, dans leur étude sur l'effet du *feedback*, Heslin et Latham (2004) ont montré que le sentiment d'auto-efficacité du manager (qui peut être mis en parallèle avec l'auto-évaluation du manager) est un modérateur de la performance. Ce sont donc les managers initialement évalués le plus favorablement qui détiennent le sentiment d'auto-efficacité le plus élevé et ce sont eux aussi qui réalisent le plus de progrès après le *feedback*.

L'accord entre son évaluation et celle des autres serait également un facteur favorisant le changement. Ainsi, pour Johnson et Ferstl (1999) les sujets qui s'étaient sur-évalués comparativement aux notations de leurs subordonnés tendent à progresser dans le temps. Ceux en accord avec les subordonnés restent stables dans leurs performances dans le temps, enfin ceux qui s'étaient sous-évalués comparativement aux notations de leurs subordonnés tendent à décroître dans leurs performances.

2. Les effets du feedback sur la performance

Un certain nombre d'études qui se sont intéressées à l'effet des évaluations à 360° et des *feedbacks* sur la performance ont montré les effets positifs des *feedbacks* sur la performance du manager à travers l'étude de variables d'effet (*outcome*) (Tornow, 1993). Ainsi, Radman (1964) a donné à des étudiants en fin de cursus une évaluation faite par leurs collègues. Deux années plus tard, la performance (le parcours et la réussite) des étudiants a été évaluée par des experts. Le résultat a montré que les étudiants les mieux notés par leurs pairs étaient ceux qui avaient eu la meilleure progression de carrière. Dans une autre étude, Edward et Evan (1996) ont étudié le résultat, dans une entreprise de production, d'une évaluation à 360° sur un certain nombre de variables d'effet (productivité, réduction des erreurs, diminution des pertes, réduction des accidents). Après un an, une hausse significative de la productivité (ne pouvant s'expliquer par d'autres

facteurs) a été constatée. De même, dans le secteur bancaire, les auteurs rapportent une progression constante de la satisfaction client (évaluée à trois reprises tous les six mois) après la réception d'un *feedback* client.

Enfin, menée en 1994 (Edward, Bland, & Kuhi, 1994) à l'Université de l'Arizona, une progression significative de la performance (six variables d'effet sélectionnées : nombre de publications référencées, nombre de publications soumises, nombre d'articles écrits, quantité de subventions proposées, reçues et de présentations aux colloques) a été notée suite à une évaluation en sources multiples (360°) comparativement aux traditionnelles évaluations *Top-down* qui existaient jusque-là.

Ces résultats ne font pourtant pas l'unanimité, un autre groupe d'études menées dans les mêmes années, ont montré que les personnes présentant une performance faible à moyenne, tendent à avoir des progressions importantes, alors qu'une baisse de la performance est observée pour le groupe des personnes très performantes (Reilly, Smither, & Vasilopoulos, 1996; Smither et al., 1995; Walker & Smither, 1999). Les mesures de la performance ont été étendues jusqu'à cinq ans et le résultat n'était pas altéré par le nombre de *feedbacks* reçus (qui ne semble pas avoir d'effet - Smither et al., 1995). Enfin, ces résultats – en tout cas une partie d'entre eux – ne peuvent être expliqués par une simple régression vers la moyenne. Par ailleurs, les auteurs de l'équipe de Smither notent que les améliorations observées l'année suivant le *feedback* sont d'autant plus importantes que les managers ont pu, l'année de l'évaluation, discuter avec leurs subordonnées des résultats obtenus.

3. Les formations

D'après les résultats précédents on pourrait être tenté de croire qu'il suffit de donner au manager des informations obtenues lors de l'évaluation à 360° pour que des effets positifs se produisent, or ce n'est pas le cas. Il est important également de donner à l'évalué des outils et des moyens pour s'améliorer afin que la diminution des différences ne soit pas réalisée en baissant l'auto-évaluation.

En effet, comme le notent Luthans et Peterson (2003) l'important n'est pas de diminuer l'autoévaluation (en créant chez le manager un doute sur ces compétences) mais de

modifier son comportement de sorte à provoquer une augmentation de l'évaluation des observateurs.

Pour ce faire, la formation doit permettre d'accroître la conscience de soi du manager car en dépit de ses bonnes intentions, la personne évaluée ne sait pas toujours comment faire pour changer. En effet, la simple mise en lumière des comportements à travailler ou des zones à problèmes ne suffit pas. La formation devient donc un élément indispensable pour aider les managers en leur donnant les clefs du changement.

Ainsi, dans une étude longitudinale, Bailey et Fletcher (2002) ont observé une augmentation significative de l'accord entre les sels et leurs supérieurs ainsi qu'entre les sels et leurs subordonnés suite à une prise de conscience par le self de ses compétences réelles qu'il avait tendance à sur-évaluer comparativement aux évaluations de ses subordonnés et à sous-estimer par rapport à celles de son supérieur.

3. Accord entre les évaluateurs

La majorité des recherches mettent en évidence de grandes différences entre l'auto-évaluation et l'évaluation des observateurs (Campbell & Lee, 1988).

James Demaree et Wolf (1993) décrivent deux aspects de l'accord entre évaluateurs:

- ✦ L'accord peut être compris comme l'étendue de la similarité sur le classement.

Cela sous-entend une approche corrélacionnelle et si la corrélation est forte, cela reflètera un modèle commun (Kenny, 1991).

- ✦ Également, l'accord peut être compris comme la différence au niveau moyen des évaluations de chacun.

Cela sous-entend une approche par ANOVA ou MANOVA (Mount, 1984). Cette approche s'intéressera donc aux différences au niveau des standards de performance.

Par ailleurs, la plupart des études portant sur les accords – désaccords entre évaluateurs accordent une place importante aux liens entre l'accord et différentes variables de réussite (variables d'effet évoquées dans la partie précédente telles que la satisfaction, le revenu, la production...)

A. L'accord au sein du groupe d'observateurs

Des variables liés à la personne, à la situation d'évaluation, aux évaluateurs ou au groupe peuvent influencer l'accord au sein des groupes d'observateurs. Selon Brutus (Brutus, Fleenor, et McCauley, 1999), cette influence peut s'exercer de deux manières différentes selon :

- ✦ Une caractéristique spécifique du self peut avoir une influence sur l'auto-évaluation du self.
- ✦ Cette même caractéristique peut influencer l'évaluation faite par les observateurs.

Même au sein d'un groupe d'observateurs (tout comme entre les groupes d'ailleurs) trois grandes explications ont pu être proposées pour expliquer les différences observées.

Tout d'abord, les évaluateurs peuvent avoir différentes opportunités pour observer les comportements de la cible (même si cet argument semble davantage s'appliquer au désaccords entre les groupes, il reste tout de même valable au sein d'un groupe de collègues ou de subordonnés).

Ensuite, les évaluateurs peuvent porter une attention plus sélective à certains comportements au détriment des autres.

Enfin, ces mêmes évaluateurs peuvent accorder une importance différente aux comportements observés (Borman, 1997; LeBreton et al., 2003; Tsui & Ohlott, 1988).

B. L'auto-évaluation

Concernant la justesse de l'auto-évaluation, le principal biais observé concerne « l'erreur fondamentale d'attribution ». Selon ce concept, nous aurions tendance à donner des explications différentes pour les réussites et pour les échecs selon qu'ils nous concernent ou bien qu'ils concernent les autres. Ainsi, on expliquerait l'absence de performance des autres plus en termes internes (il / elle échoue du fait de ses faibles capacités) et notre manque de performance en termes externes (j'échoue du fait des circonstances).

De plus, pour conserver le plus possible une bonne image de nous-même et d'après la théorie de l'optimisme comparatif (Weinstein, 1980), nous aurions tendance à nous auto-évaluer plus favorablement que l'on n'évaluerait autrui.

Ainsi, en compare d'un côté l'auto-évaluation réalisée par le self et de l'autre la vision qu'a le self de la perception des autres sur lui-même, nous obtenons des taux d'accord moyens à élevés (Shrauger & Schoeneman, 1979). Par contre l'évaluation de ce même self n'est pas un bon prédicteur de l'évaluation réelle faite par les observateurs (Mabe & West, 1982).

C. L'accord entre les groupes d'observateurs

Un certain nombre de théories se proposent d'expliquer de manière globale, les désaccords observés entre les groupes.

La présence ou absence d'homogénéité (*Dissimilarity Effect* – Roberson & Block, 2001) aurait une influence sur l'accord. Plus exactement, l'absence d'homogénéité entraînerait une sous évaluation.

La théorie de la similarité et de l'attraction (Riordan, 2000) explique que les individus s'identifient et sont attirés par ceux qui leur sont similaires, ce qui entraîne une réponse attitudinale plus positive dans le cas de similarité. Aussi, dans les groupes d'observateurs, quand les individus sont similaires les uns aux autres (âge, sexe, race..) les variables d'effet tendent à être plus élevées.

Les stéréotypes et les attentes des évaluateurs jouent également un grand rôle. Si le supérieur évalué n'est pas un homme blanc (stéréotype du manager) son évaluation par les observateurs sera plus faible car il est dissimilaire à l'attente selon la théorie des stéréotypes et de l'exagération des différences (Ilgen & Youtz, 1989).

La théorie de l'attribution (Jones & Nisbett, 1972; Lord et al., 1986) prévoit, quant à elle, que les selves auraient tendance à attribuer les bonnes performances à leur comportement et les mauvaises à l'environnement, alors que les observateurs attribueraient plus volontiers les bonnes performances à l'environnement et les mauvaises à l'acteur (selves). Cela créerait donc des désaccords constants entre les selves et les observateurs.

Enfin, suite aux travaux d'Ashford (1989), Atwater et Yammarino (1997) ont proposé un modèle global pour expliquer le désaccord. D'après ces auteurs, un ensemble de caractéristiques telles que l'expérience, les informations biographiques ou la personnalité affectent la perception de soi des individus et donc, par effet de conséquence, affecte leur auto-évaluation.

Dans leur modèle, les leaders dont l'auto-évaluation est élevée et en accord avec celle des observateurs auront des corrélations élevées avec des variables d'effet positives (au niveau individuel et organisationnel). Par contre, les leaders qui sont en accord avec les observateurs sur leur propre faible performance auront, eux, des corrélations plus importantes avec des variables d'effet négatives.

D. Variables pouvant influencer l'accord entre observateurs

Nous avons vu précédemment les théories proposant des explications d'ensemble concernant l'accord et le désaccord, nous allons aborder maintenant des variables plus spécifiques, se proposant d'expliquer, elles aussi, des différences observées dans les évaluations à 360°.

1. Le genre

D'après Daubman et ses collaborateurs (1992), les femmes auraient en général une plus faible confiance en elles dans un contexte de réussite. Cela entraînerait donc une sous-évaluation de leurs capacités comparativement aux évaluations des hommes. Cependant, cette sous-évaluation s'estomperait, voire disparaîtrait, dans le cas où l'évaluation serait faite de manière anonyme. Les auteurs en concluent donc que l'auto-évaluation des femmes serait plus sensible aux variables contextuelles.

Les hommes comparativement aux femmes, seraient moins réceptifs aux *feedbacks* (Fletcher, 1999) et seraient moins enclin à modifier les perceptions que les autres ont d'eux. De ce fait, ils auraient tendance à se sur-évaluer mais aussi à sur-estimer leur niveau d'efficacité (Brutus et al., 1999; Jones & Fletcher, 2002; Vecchio & Anderson, 2009). De la même manière, les hommes auraient tendance à sur-estimer leurs aptitudes dans le domaine spatial, mathématique et kinesthésique (Visser, Ashton, & Vernon, 2008) mais aussi leur niveau de leadership transformationnel (Moshavi, Brown, & Dodd, 2003) comparativement aux femmes.

En ce qui concerne les évaluations réalisées par les observateurs, les femmes seraient évaluées plus faiblement que les hommes pour les tâches considérées comme masculines dont le management est un très bon exemple (Brutus, Montei, Jex, King, & King, 1993).

Ainsi dans le cadre des 360°, le genre semble être un facteur permettant d'expliquer les différences observées entre les évaluations des sels et celles des observateurs.

2. L'ethnicité

Les groupes minoritaires sont souvent sujets aux stéréotypes négatifs particulièrement lorsqu'un individu issu de ces groupes occupe un poste de direction et, contrairement aux femmes (elles aussi minoritaires dans le contexte de l'entreprise) les préjugés portant sur les minorités ethniques ne diminuent pas avec la hausse des échelons hiérarchiques (Nielsen, Fernandez, & Peng, 1982).

Les managers issus de groupes minoritaires sont souvent sujets à des attentes plus faibles ou moins marquées ce qui est susceptible d'entraîner des évaluations altérées de la part des observateurs. Par ailleurs, le *feedback* reçu par ces managers issus de groupes minoritaires tend souvent à être « dysfonctionnel » ce qui peut créer des inadéquations dans leurs auto-évaluations (Coleman, Jussim, & Isaak, 1991) et une sur-estimation de leurs performances (Larwood, 1982; Ostroff, Atwater, & Feinberg, 2004).

3. L'âge

En ce qui concerne l'âge, en vieillissant, les individus tendent à se voir de manière plus positive (Brutus et al., 1999) et plus efficace (Ostroff et al., 2004; Vecchio & Anderson, 2009). En même temps, ils deviennent moins réceptifs aux *feedbacks* ce qui crée une double sur-évaluation lorsqu'ils s'auto-évaluent.

Les observateurs, quant à eux, tendent à avoir une perception plus négative des employés âgés (stéréotype) ce qui entraîne une sous-évaluation de leurs qualités par les observateurs (Ostroff et al., 2004).

Les managers les plus âgés tendent donc à accroître leur auto-évaluation (Cleveland & Landy, 1987; Wohlers & London, 1989) et leurs subordonnés tendent à décroître leurs notations (Cleveland & Landy, 1987; Tsui & O'Reilly, 1989) d'où des différences observés plus importantes avec l'âge du manager évalué.

4. L'expérience

Les personnes plus expérimentées tendent à se voir de manière plus positive que les moins expérimentées (Brutus et al., 1999). Elles recherchent également moins le *feedback* et, qui plus est, dû fait de leur position hiérarchique (avec l'expérience ils sont souvent à des postes élevés) elles en reçoivent de moins en moins souvent (Gentry, Yip, & Hannum, 2010). Ces deux éléments créent une forte tendance à la sur-évaluation et entraînent un « décalage » plus marqué avec les observateurs (Sala, 2003).

5. L'éducation

Les personnes plus éduquées auraient des aptitudes analytiques plus développées et une plus grande complexité cognitive (Kingston, Hubbard, Lapp, Schroeder, & Wilson, 2003). Elles feraient donc un traitement plus efficace de l'information et seraient plus à même d'utiliser efficacement le *feedback*. De ce fait, elles seraient plus capables d'adapter leur auto-évaluation aux évaluations des observateurs. Cela créerait donc plus d'accord pour les managers plus éduqués (Ostroff et al., 2004) comparativement aux moins éduqués.

6. La Similarité self - observateur

Bien que les variables que nous avons évoquées jusque-là aient des effets sur l'évaluation que des selves et des observateurs, il s'agit souvent d'un effet d'interaction. À présent, nous allons nous intéresser aux variables de similarité et de différences entre les selves et les observateurs, et à leurs effets sur les évaluations.

A. Le genre et la composition du groupe

Comme nous l'avons vu, dans leurs auto-évaluations, les femmes tendent à se sous-évaluer comparativement aux hommes. De plus, si le groupe est composé de personnes similaires (féminin) l'évaluation des observateurs sera plus favorable que s'il est dissimilaire (Sackett, DuBois, & Noe, 1991).

Ainsi, les femmes qui supervisent surtout des femmes seront jugées plus favorablement par les observateurs mais, comme elles ont tendance à se sous-évaluer, cela entraînera

un désaccord. Inversement, les femmes qui supervisent surtout des hommes seront jugées moins favorablement par les observateurs et comme elles aussi, ont tendance à se sous-évaluer cela entraînera paradoxalement un accord entre l'auto-évaluation de la femme manager avec l'évaluation des observateurs.

Enfin, les hommes supervisant surtout des femmes seront notés moins favorablement et comme les hommes tendent à se sur-évaluer cela entraînera donc un grand désaccord. Ostoff et al., (2004) ont trouvé par ailleurs qu'une similarité de genre et d'âge des managers avec leurs évaluateurs (subordonnés, pairs et superviseurs) n'était pas reliée directement à l'auto-évaluation du manager.

B. L'âge et la composition du groupe

Nous avons vu que les managers plus âgés tendent à se surévaluer. De plus l'influence d'un stéréotype concernant les employés âgés fait qu'ils tendent à être sous-évalués comparativement aux employés jeunes, introduisant donc une grande différence entre l'auto-évaluation et l'évaluation des observateurs.

Mais, si le manager travaille avec des employés plus âgés que lui (dissimilarité) il peut ressentir une sensation d'insécurité et d'incompétence (même processus que celui à l'œuvre pour les femmes manager, exerçant dans un environnement majoritairement masculin) entraînant donc une sous-évaluation. Ce qui, au final, permet de réduire la différence entre les évaluations.

C. Nationalité et la composition du groupe

Dans une étude récente, Caligiuri et Day (2000) ont montré que la similarité de nationalité peut, elle aussi, avoir un effet sur l'accord. En effet, en ce qui concerne les évaluations des observateurs, lorsque les managers sont de la même origine qu'eux, ces derniers tendent à les noter plus favorablement.

7. Variables de personnalité / aptitude

A. L'intelligence

Les aptitudes cognitives ont un effet non négligeable sur les évaluations dans le cadre des 360°. Ainsi dans leur méta analyse, Mabe et West (1982) citent de nombreuses études appuyant l'idée que plus un individu est intelligent, plus son auto-évaluation tend à être similaire aux évaluations objectives de performance ou de réussite.

De même, Yammarino et Atwater (2001), expliquent que plus une personne est intelligente et plus elle possède une plus grande complexité cognitive, meilleur sera son souvenir. De ce fait, elle pourra mieux stocker et analyser l'information qui lui sera communiquée et elle le faire avec une plus grande précision. Par conséquent, elle aura une évaluation plus exacte d'elle-même et en recevra une plus similaire de la part des autres (Beehr, Ivanitskaya, Hansen, Erofeev, & Gudanowski, 2001; London, 1995).

B. La conscience de soi

Pour London et Smither (1995) la conscience de soi serait le modérateur de la relation entre le *feedback* du 360 et la hausse de la performance. Les managers les plus performants auraient une plus grande conscience d'eux-mêmes comparativement aux moins performants (Atwater & Yammarino, 1992; Bass & Yammarino, 1991) et de ce fait, le niveau de divergence entre leur auto-évaluation et l'évaluation des observateurs devrait diminuer. Cela expliquerait que les managers plus performants, grâce à l'intervention de la conscience de soi, auraient des divergences moindres que les managers moins performants.

C. La confiance en soi

Les individus présentant une forte confiance en eux auraient tendance à se sur-estimer comparativement aux individus ayant une faible confiance en eux (Harris & Schaubroeck, 1988), ils devraient donc présenter des divergences plus importantes avec les évaluations de leurs observateurs.

D. Le locus de contrôle

D'après Seeman (1963; 1962) une perception de contrôle interne (Locus de contrôle interne) serait liée à une capacité plus élevée à rechercher, retenir et intégrer des informations portant sur notre personnalité et nos habilités (Mabe & West, 1982) De ce fait, un locus de contrôle interne serait lié à une auto-évaluation plus adéquate.

Or, d'autres auteurs ont montré que les managers avec un locus de contrôle interne présentaient une sur-évaluation de leur potentiel managérial (Levy, 1993), ce qu'ils interprètent en référence à la théorie de l'attribution : ceux qui se considèrent responsables et à l'origine de leurs réussites, ont tendance, par là même, à juger leur comportements plus positivement.

Cette interprétation n'est pourtant pas unanime, Whetten et Cameron (2007) proposent de ré-interpréter ce résultat en disant que ceux ayant un locus de contrôle interne sont peut-être simplement de meilleurs managers et de ce fait, il ne s'agirait pas d'une sur-évaluation, ce qui rejoindrait la conclusion de Seeman.

E. Assertivité

D'après Atwater et ses collègues, (2009) l'accord entre le self et ses subordonnées, de même que l'accord entre le self et ses collaborateurs serait plus élevé dans les pays caractérisés par un haut niveau d'assertivité.

Pour les auteurs, les cultures occidentales sont caractérisées par un fort niveau d'assertivité, la communication y est directe et l'interlocuteur s'attend à une expression des idées non ambiguës. Ces cultures valorisent également le partage d'opinions, d'idées et de sentiments. Contrairement à elles, les cultures moins assertives préfèrent l'utilisation d'un langage moins direct et plus subtil et valorisent le détachement émotionnel et la discrétion. La classification des pays sur cette dimension est issue de la base donnée du projet GLOBE (House, Hanges, Javidan, Dorfman, & Gupta, 2004).

Atwater et ses collègues ont ainsi trouvé que l'accord entre les selves et leurs subordonnés ainsi qu'entre les selves et leurs collaborateurs sur les dimensions du *Benchmarks* (Center for Creative Leadership, 2004) était plus élevé dans les pays occidentaux.

F. Dominance

En utilisant l'échelle CPI (*California Personality Inventory*) Brutus et ses collaborateurs (1999) ont pu prédire le désaccord entre les observateurs et le selves. En effet, plus les managers avaient des scores élevés sur cette dimension, plus ils tendaient à sur-estimer leur performance. De même, une étude plus récente (Jackson et al., 2007) a montrée que le niveau de dominance était lié avec une sur-estimation de la performance.

D. Les catégories d'observateurs

En plus de ces différences observées en général entre les auto-évaluations et les évaluations par les observateurs, certains auteurs se sont penchés sur la spécificité des divers groupes d'observateurs évaluant le self (Brutus et al., 1999).

Il a pu être montré que les managers n'utilisaient pas les mêmes tactiques pour influencer leurs supérieurs que leurs subordonnées (Yukl & Tracey, 1992). Les perceptions de ceux-ci ne sont donc pas forcément les mêmes. Yukl et Tracey notent également que tant qu'à choisir, un manager aura tendance à faire de son mieux pour correspondre à la vision de son supérieur comparativement à la vision de ces pairs ou subordonnés – et ce d'autant plus que ces visions sont différentes.

Pourtant, dans leur étude Brutus et ses collaborateurs (1999) ont montré que les facteurs qui influencent et prédisent le désaccord entre les selves et leurs pairs sont les mêmes que ceux qui prédisent le désaccord des selves avec leurs subordonnés.

Par contre, dans l'ensemble, les caractéristiques démographiques ou de personnalité sont le moins à même d'expliquer la variance des observations du supérieur, comparativement aux variances des autres catégories d'observateurs.

Nous avons donc vu dans la partie précédente l'importance donnée aux écarts et aux différences observées au sein des évaluations à 360°. Beaucoup de variables explicatives se proposent de prédire et d'expliquer leur existence, mais qu'en est-il sur le plan statistique ? Comment les évaluations à 360° parviennent-elles à passer outre les difficultés posées par ces divergences d'opinion ?

4. Traitement des scores et évaluation des écarts

Étant donné l'organisation des données recueillies dans le cadre des analyses à 360° : plan à mesure répétées ; organisation hiérarchique de la structure ; emboîtement observateurs / manager etc. des techniques d'analyse spécifiques sont nécessaires. Pourtant, elles n'ont pas toujours été utilisées et les études en faisant l'économie n'ont que trop rarement estimé les biais liés à l'utilisation de méthodologies ne prenant pas en compte ces spécificités des données.

A. Analyses multi-niveaux

L'une des difficultés majeures des données issues des questionnaires à 360° est leur inter-dépendance. En effet, pour un même sujet, plusieurs observateurs donnent leurs évaluations, celles-ci ne sont donc pas indépendantes.

De plus, comme signalé à de nombreuses reprises par Edwards (1994, 1995, 2002) et Edwards et Parry (1993), depuis le début des recherches sur le SOA (*Self – Other Agreement*) un grand nombre d'études ont adopté une approche classique dans leur traitement des données des questionnaires à 360°, consistant à faire des moyennes par catégories d'observateurs. Puis dans un second temps, cette moyenne était utilisée pour calculer un indice de différence entre les évaluations des observateurs et l'évaluation des selves. Or cette méthode donne souvent des conclusions ambiguës et, qui plus est, présente la plupart du temps une fidélité plus faible que celle de chacun de ses composants pris individuellement.

D'après Chan (1998), les modèles multi-niveaux peuvent être organisés en cinq catégories selon les relations fonctionnelles que les phénomènes entretiennent aux différents niveaux d'analyse. Cinq catégories permettent d'organiser les modèles en fonction de la définition donnée au construit supérieur au regard du construit de niveau inférieur : le modèle additif, le modèle de consensus direct, le modèle de consensus avec référent, le modèle de dispersion et le modèle de processus.

Dans le modèle Additif, les construits des différents niveaux entretiennent des relations fonctionnelles directes. Le score de niveau supérieur est donc directement calculé comme la somme des scores des niveaux inférieurs et ce, sans tenir compte de la variance de ceux-ci. En effet, dans ce modèle, la variance (que ce soit celle des niveaux inférieurs ou

celle des niveaux supérieurs) n'est d'aucun intérêt ni théorique, ni conceptuel. Donc pour obtenir le score du niveau supérieur, par exemple de groupe, le chercheur ayant recours au modèle additif, fait la moyenne des scores individuels sans tenir compte de la variance inter-sujets.

Or, dans les recherches portant sur des résultats d'évaluations à 360°, la dispersion présente de l'intérêt comme en témoignent les nombreuses études sur le désaccord entre les catégories d'observateurs. Donc, dans ce cas le modèle Additif de Chan (1998) n'est plus le modèle adéquat. Pour les résultats de questionnaires à évaluateurs multiples, le modèle de « Consensus direct » devient donc le modèle approprié.

Le modèle de Consensus Direct utilise le consensus intra-groupe au niveau inférieur pour justifier la relation fonctionnelle avec le construit supérieur et donc pour l'opérationnaliser. Ce modèle postule donc que les contenus opérationnalisés à un niveau inférieur sont fonctionnellement isomorphes à une autre forme de construit de niveau supérieur. Un indice d'accord intra-groupe tels que l'indice r_{wg} (James, Demaree, & Wolf, 1984) ou le Coefficient de Corrélation Intraclasse (Shrout & Fleiss, 1979) peut-être calculé afin de justifier l'agrégation des réponses individuelles.

Ainsi, si nous voulons agréger des données de niveau inférieur (par exemple les évaluations individuelle des observateurs de type « Subordonnés ») en un score de groupe de niveau supérieur (par exemple le score « Subordonné ») nous devons nous pencher sur la similarité des construits d'ordre inférieur (Kozlowski & Klein, 2000).

Pour éprouver cette similarité, deux démarches peuvent être envisagées : l'approche par consistance ou l'approche par consensus.

Les **approches par consistance** ou par fidélité, examinent la fidélité inter-juges pour chaque dimension sur l'ensemble des groupes en utilisant des indices statistiques comme les Coefficients de corrélation Intraclasse (ICC – *Intraclass Correlation Coefficient*) ou les analyses ANOVA (*ANalysis Of VAriance*) qui comparent les variances Inter- et Intra-groupes au regard de la variance totale. Il s'agit donc d'approches par échantillon (Kozlowski & Hattrup, 1992).

Les **approches par consensus** ou par accord, se proposent de comparer la variance intra-groupe avec une variance hypothétique qui estimerait le hasard. L'accord au sein du groupe de niveau inférieur est ainsi examiné pour chaque mesure, pour chaque sous-

ensemble, avec des indices tels que le $r_{wg(J)}$. Il s'agit donc d'une approche par groupe (Kozlowski & Hattrup, 1992).

Nous voyons donc qu'au niveau théorique, le calcul et l'agrégation des scores issus des analyses à 360° n'est pas forcément évidente, et la moyenne des évaluations des observateurs n'est généralement pas une statistique appropriée pour englober les proportions d'unités configurales (Brown, Kozlowski, & Hattrup, 1996).

B. Similarités inter- et intra-groupes

Du fait de son objectif premier, le questionnaire à 360° vise à apporter au self des évaluations de lui-même provenant d'autres sources. Il entend donc mettre en relation la perception que le self a de lui-même (d'elle-même) avec la perception que d'autres ont de lui (elle). Pour ce faire, ces autres (les observateurs) sont regroupés selon leur rapport hiérarchique avec le self : supérieur, subordonné, collègue... Aussi, la majorité des recherches acceptent implicitement l'hypothèse des écarts

Cette hypothèse veut que les évaluations au sein des groupes d'observateurs (pairs, supérieurs...) soient fortement similaires, mais que ces évaluations soient faiblement similaires entre les groupes d'observateurs (pairs – subordonnés ; subordonnés – supérieur...).

Un grand nombre de recherches ont pu éprouver cette hypothèse en obtenant de manière stable des coefficients de corrélations de Pearson ou des Coefficients de corrélation Intraclasse modérés au sein des catégories d'observateurs et des coefficients plus faibles entre les catégories d'observateurs (Borman, 1997; Conway & Huffcutt, 1997; Viswesvaran, Schmidt, et Ones, 2002).

Pourtant, au début des années 2000, Le Breton (2003; LeBreton & Senter, 2008) et ses collaborateurs ont proposé une nouvelle approche de cette question. En effet, d'après ces auteurs le niveau d'accord observé dans les études antérieures serait en grande partie induit par les indices statistiques utilisés. Dans les études précédentes, les chercheurs utilisent des indices corrélationnels pour montrer le faible lien existant entre les groupes d'observateurs de catégorie différente. Or pour Le Breton, ces résultats peuvent tout aussi bien s'expliquer par un autre élément : un problème de restriction de la variance.

Cette hypothèse de restriction de la variance (LeBreton et al., 2003; LeBreton & Tonidandel, 2008) repose sur l'idée que les managers qui sont l'objet d'évaluations à 360°,

représentent souvent un échantillon assez restreint de la population. De même, la tendance générale des observateurs à donner des évaluations plutôt élevées amène, elle aussi, à une variabilité faible dans les données. Les auteurs suggèrent donc que les corrélations faibles trouvées précédemment seraient dues aux méthodes corrélationnelles utilisées et non pas à une absence d'accord entre les groupes d'observateurs à proprement parlé.

Aussi, en utilisant des indices non sensibles à la restriction de la variance (tels que les indices $r_{wg(j)}$), Le Breton et ses collaborateurs trouvent que les accords Inter- et Intra-groupe sont du même ordre, et qu'ils sont élevés.

Il subsiste donc aujourd'hui un manque de clarté en ce qui concerne l'accord au sein et entre les groupes d'observateurs. Nous verrons dans la partie d'analyse, que des méthodes statistiques adaptées permettent de mieux comprendre ces divergences.

C. Théorie de la généralisabilité

Au vu de la structure particulière des données issues de l'approche à 360° que nous venons de décrire, le modèle psychométrique classique semble difficilement applicable.

En effet, l'équation de base du modèle classique de la mesure stipule que le score observé est composé d'un score qualifié de vrai et d'une erreur de mesure. Or cette erreur est indifférenciée, ce qui vu notre plan de mesure est difficilement concevable. Aussi, bien que leur travaux se soient développés dans le domaine de la psychologie de l'éducation, la théorie de la généralisabilité de Cardinet et Tourneur (1985) présente un grand intérêt pour nous. Alors que leur approche visait principalement à proposer un modèle adapté aux mesures éducatives, cette théorie permet également un éclairage intéressant pour comprendre les difficultés liées aux plans de mesure à 360°.

La théorie de la généralisabilité (Cardinet & Tourneur, 1985; Cardinet, Tourneur, & Allal, 1976) cherche à dépasser la conception unitaire de l'erreur de mesure en la différenciant selon les sources. Cette théorie permet donc de distinguer les sources d'erreur dans les résultats. Cela est valable pour les mesure pour lesquelles un plan à été établi à l'avance pour spécifier les sources d'erreur qui doivent être analysées. Ainsi, tous les aspects d'une situation de mesure peuvent être caractérisés : les items, les évaluateurs, le moment de

mesure... L'objectif étant de déterminer la source la plus importante pour pouvoir la contrôler.

Ainsi comparativement à l'équation de base de la théorie classique, l'équation en théorie de la généralisabilité décrit le score observé comme la somme du score univers (pendant du score vrai) avec un ensemble d'erreurs de mesure provenant de sources différentes.

Nous verrons dans la partie d'analyse, comment l'idée des sources d'erreur différentes permet de mieux définir le modèle de mesure spécifique au 360°.

D. Équivalence de la mesure

Depuis de le début des années 2000, de nombreuses recherches ont porté sur l'étude des équivalences de la mesure entre les sources. Ainsi, les travaux de Cheung (1999), mettent en lumière l'importance de vérifier cette équivalence.

D'après l'auteur deux types d'équivalence doivent être prises en compte : l'équivalence conceptuelle et l'équivalence psychométrique. La première indique que les items du test répondent à la même structure factorielle pour les divers groupes d'évaluateurs. Alors que la seconde, l'équivalence psychométrique, indique que non seulement l'instrument a la même structure factorielle dans les divers groupes, mais aussi, que les individus répondent de la même manière selon leur catégorie d'appartenance (self, collègue, supérieur...)

Beaucoup d'auteurs (Dieffendorff, Silverman, & Greguras, 2005; Fecteau & Craig, 2001; Maurer, Raju, & Collins, 1998) se sont penchés sur l'étude de ces équivalences avec pour objectif d'expliquer certains désaccords ou différences observées entre les groupes. Dans leur méta-analyse, Vandenberg et Lance (2000) rappellent l'importance de s'assurer de l'équivalence des mesures entre les groupes avant d'envisager quelconques comparaisons. Ainsi, de même que de nombreux autres avant eux (Drasgow, 1984, 1987; Horn & Mcardle, 1992; Vandenberg & Self, 1993), Vandenberg et Lance mettent en garde les chercheurs contre des comparaisons de groupe sans vérification d'équivalence à priori, que ce soit grâce aux analyses factorielles confirmatoires, au modèle de réponse à l'item ou à la théorie de la généralisabilité.

La question psychométrique prend donc une place de plus en plus grande dans l'approche des questionnaires à 360°. Du fait de leur structure particulière et de leurs spécificités, nombreux sont les raisons qui poussent à utiliser des méthodologies et des outils d'analyses pouvant tenir réellement compte de la complexité des données.

Pourtant, il est intéressant de noter qu'une grande partie des études s'intéressant aux équivalences des mesures – ne s'intéressent pas à l'homogénéité des groupes.

Ainsi, Diefendorff et ses collègues (2005) éludent le problème en « sélectionnant au hasard une seule évaluation pour les sels ayant plusieurs répondant dans certaines catégories (plus d'un collègue ou plus d'un supérieur) » (Diefendorff et al., 2005, p. 409).

De même, Hannum (2007) qui s'est intéressée aux équivalences entre groupes d'observateurs (supérieurs, collègues et subordonnés), a choisi de créer un « évaluateur hypothétique » (en moyennant les réponses par item par groupe d'observateur) et ce sans s'intéresser à l'accord intra-groupe.

Plus complexe, le plan d'analyse de Maurer et de ses collègues (1998) propose une sélection aléatoire de deux collègues et de deux subordonnés par manager évalué. Dans la suite de l'analyse, ces deux évaluations sont utilisées successivement : d'abord pour construire le modèle (Subordonné 1 et Collègue 1) puis dans une seconde phase pour tester le modèle (Subordonné 2 et Collègue 2).

L'ensemble des auteurs s'accorde donc sur la complexité des données issues des questionnaires à 360°, pourtant peu d'entre eux proposent une méthodologie permettant de tenir compte de l'ensemble de ces conséquences liées à celles-ci.

Dans le cadre de ce chapitre nous avons décrit la spécificité des évaluations à 360° en nous penchant tout d'abord sur leurs particularités puis sur les difficultés qui pouvaient y être associées. Nous nous sommes ensuite intéressés aux diverses variables mises en lumière au cours de plusieurs décennies d'étude sur les évaluations multi-source et permettant d'expliquer ou de prédire l'existence de différences entre les sources d'évaluations (entre les subordonnés, les pairs, les supérieurs) ou entre l'évaluation du self et l'évaluation des observateurs.

Dans la seconde partie de ce chapitre, nous avons constaté la spécificité statistique des données issues des analyses à 360° et des nécessaires aménagements que leur analyse requiert. Il apparaît clairement à présent que l'accord entre les observateurs est un point central des études à 360° et que celui-ci doit être abordé avec les outils statistiques adéquats et en tenant compte des structures de données associées avec les 360°.

Nous allons donc dans notre démarche d'analyse nous intéresser, nous aussi, à l'accord au sein des observateurs, en l'abordant d'abord avec la méthode multi-niveaux classique puis en discutant l'hypothèse de la restriction de la variance telle qu'énoncée par LeBreton. Nous montrerons ainsi les intérêts mais aussi les limites de cette approche. Ensuite, dépassant l'approche par groupe et par échantillon, nous chercherons à mieux comprendre les écarts observés dans les évaluations des observateurs d'un même self. Certaines variables permettent-elles de mieux comprendre ces divergences ? Des caractéristiques du self ou des observateurs sont-elles à même d'expliquer cette variabilité des évaluations ? Enfin, nous prendrons de la hauteur et chercherons à mieux comprendre la structure même du 360 et la manière avec laquelle la théorie psychométrique peut s'y retrouver. Nous chercherons donc à proposer une compréhension psychométrique et théorique des scores et des différences obtenues dans le cadre des évaluations à 360°.

Méthode

Dans cette partie nous allons présenter l'outil sur lequel portera l'ensemble de nos analyses – le GELI. Nous présenteront également notre échantillon et ses caractéristiques.

1. L'outil de mesure : le *Global Executive Life Inventory (GELI)*

Le *Global Executive Leadership Inventory* (Kets de Vries et al., 2004) est un outil complet d'évaluation du leadership que nous avons étudié en détail dans le chapitre 3 (Chapitre 3.4.B.4). Il s'inscrit dans l'approche clinique du leadership et fait partie des instruments de mesure proposant une évaluation en 360°, pouvant donc être complété aussi bien par le manager évalué que par des observateurs de celui-ci et offrant des consignes appropriées pour chacun de ces types de répondants.

A. Les dimensions du GELI

Comme nous venons de le rappeler, le GELI (Kets de Vries et al., 2004) est donc un instrument développé pour mesurer les traits propres à l'approche clinique du leadership. Il évalue 12 caractéristiques du leader et permet une évaluation à 360° afin de confronter la perception que le leader a de ses qualités avec la perception qu'en ont ses collaborateurs, supérieurs et subordonnés.

Ces 12 échelles sont composées de 5 à 12 questions :

- *Envisioning,*
- *Empowering,*
- *Energizing,*
- *Designing and aligning,*
- *Rewarding and feedback,*
- *Teambuilding,*
- *Outside orientation,*
- *Global mindset,*
- *Tenacity,*
- *Emotional intelligence,*
- *Life balance*
- *Resilience to stress.*

Les deux dernières échelles *Life Balance* et *Resilience to stress* n'ont pas été analysées dans le travail ci-après car, comme elles portent sur des comportements qui ne peuvent pas toujours être observés au travail, on a permis aux observateurs de ne pas répondre à la question s'ils pensent qu'ils n'ont pas les informations pour répondre de manière pertinente. Or, au vu des démarches statistiques que nous allons entreprendre dans la suite de ce travail, la présence de non-réponses systématiques ne peut être gérée par des solutions adaptées particulièrement pour les méthodes multi-niveaux. Rappelons que les travaux de Rubin (1987) proposent une typologie des non-réponses : complètement au hasard (*completely at random*), au hasard (*at random*) et ne pouvant être ignorées (*not ignorable*). Les réponses manquantes « ne pouvant être ignorées » qui relèvent d'un processus systématique (ici la relation entre l'observateur et le self qui ne lui a pas permis d'observer le comportement) ne peuvent faire l'objet de procédures classiques de remplacement ou d'imputation. Pour cette raison, nous avons préféré exclure ces deux variables des analyses.

Nous avons donc fait porter nos analyses sur les 10 premières échelles du test, qui correspondent à 80 propositions à coter sur une échelle de type Lickert en 7 points.

Le point 1 est présenté comme « Ne me correspond pas du tout » (« *Not at all like me* ») ou « Je n'agis jamais de la sorte » (« *I never act this way* »), pour la version Self - lorsque le sujet répond au questionnaire sur ses propres qualités, et comme « Ne lui correspond pas du tout » (« *Not at all like him/her* ») ou « Il/Elle n'agit jamais de la sorte » (« *He/She never acts this way* ») pour la version Observateur - lorsque les subordonnés, collaborateurs ou supérieurs répondent à propos du manager évalué en 360.

De même, le point 7 est présenté comme « Me correspond totalement » (« *Very much like me* ») ou « j'agis toujours de la sorte » (« *I always act this way* »), pour la version Self, et comme « Lui correspond totalement » (« *Very much like him/her* ») ou « Il/Elle agit toujours de la sorte » (« *He/She always acts this way* ») pour la version Observateur.

L'outil a été dans un premier temps développé et validé en anglais. Aujourd'hui des versions en d'autres langues existent, tel que le français et le russe par exemple. Dans le cadre de nos analyses nous n'utiliserons que les données de la version anglaise.

B. Administration et passations

Le GELI a été administré lors des séminaires et des programmes de formation mis en place par l'INSEAD et plus spécifiquement lors du programme dirigé par le Pr. Manfred Kets de Vries « *The Challenge of Leadership: Creating Reflective Leader* » au sein du laboratoire de recherches international *Global Leadership Center*.

Ce programme vise à former les managers en exercice, afin de les aider à comprendre leur faiblesses et leurs forces dans le domaine du leadership, tout en les accompagnant dans le développement de celles-ci.

La passation du GELI se déroulait en début de formation et visait à donner au manager un état des lieux sur, d'une part, sa perception de ses qualités de leader et, d'autre part, sur la perception qu'ont ses collègues, ses subordonnés, ses supérieurs de ses qualités.

Les managers proviennent de grandes sociétés internationales et multi-nationales, ils exercent à des postes de responsabilité et sont pour la plupart en milieu de carrière. En amont de la formation, le manager devait choisir les personnes qui auraient le rôle d'observateur et donc évalueraient ses qualités de leader.

La passation du questionnaire GELI se réalise sur ordinateur, sur internet. Le manager reçoit un e-mail de l'INSEAD l'invitant à compléter une évaluation en ligne. Pour ce faire, l'e-mail comporte un lien ainsi que des codes lui permettant d'accéder au site internet sécurisé où il peut répondre aux questions du test.

En plus de quelques informations biographiques (âge, genre, poste occupé, secteur d'activité de l'entreprise...) le manager doit renseigner le nom et l'adresse e-mail de ses observateurs. Suite à cela les observateurs reçoivent, à leur tour, un e-mail les invitant à compléter un questionnaire sur le manager en question.

La passation du test se déroule de manière similaire pour le manager et ses observateurs. Après la présentation de la consigne, les items apparaissent sur l'écran de l'ordinateur et la personne répondant au questionnaire doit sélectionner sur une échelle de Lickert en cinq points sa réponse pour chacune des propositions.

Le logiciel informe également les observateurs que leurs réponses seront anonymisées à l'exception de celles du supérieur. Cela signifie que seule la catégorie d'observateur sera utilisée lors de la restitution des résultats, et ceux-ci seront analysés en moyenne, si une autre catégorie que la catégorie Supérieur venait à présenter une seule observation; celle-ci serait transférée dans la catégorie Autre.

En effet, lors du débriefing et de la séance de restitution des résultats avec le manager, le formateur lui présentera les opinions de chacun des groupes d'observateurs dans leur ensemble et non pas les évaluations individuelles de chaque observateur. Ainsi, le manager saura, qu'en moyenne, sur telle ou telle autre dimension, ses collaborateurs ont qualifié sa compétence comme faible, moyenne ou forte, mais il n'aura pas d'information sur l'évaluation spécifique de tel ou tel autre observateur.

De même, les observateurs sont informés que leurs évaluations ne seront utilisées que pour des actions de formation et de développement personnel et qu'en tant qu'observateurs, ils n'auront pas accès aux résultats de l'évaluation.

C. Qualités psychométriques

Le GELI présente de bonnes qualités psychométriques, avec des alphas de Cronbach allant de 0.76 à 0,91 pour chacune des dimensions Tableau A.1 (Kets de Vries et al., 2004).

Dimension	Nombre d'items	Alpha (Self)*	Alpha (Observateur)**
<i>Visioning</i>	8	0,73	0,79
<i>Empowering</i>	8	0,72	0,82
<i>Energizing</i>	8	0,76	0,84
<i>Designing & Aligning</i>	7	0,81	0,87
<i>Rewarding & Feedback</i>	8	0,82	0,89
<i>Teambuilding</i>	11	0,84	0,88
<i>Outside Orientation</i>	5	0,77	0,86
<i>Global Mindset</i>	8	0,87	0,88
<i>Tenacity</i>	5	0,73	0,79
<i>Emotional Intelligence</i>	12	0,87	0,93

Tableau A.1 Nombre d'items et coefficients alphas de Chronbach pour les 10 échelles du GELI.

* n=320 ** n=287

Kets de Vries M., F., R., Vrignaud, P., Florent-Treacy E. (2004) The Global Leadership Life Inventory: development and psychometric properties of a 360-degree feedback instrument. *International Journal of Human Resource Management*, 15, p.484

2. L'échantillon

Les données utilisées dans cette thèse proviennent des évaluations faites avec le *Global Executive Life Inventory* (GELI) entre 2003 et 2006.

Durant ces trois années, de nombreuses personnes ont complété ce questionnaire. Ainsi, la même base a été utilisée pour enregistrer les résultats des évaluations réalisées dans le cadre de programmes Intra-entreprises, mais aussi, de programmes destinés aux étudiants en MBA. Afin de s'assurer que les résultats ne portent que sur des managers en exercice, les évaluations effectuées auprès des étudiants en MBA ont été enlevées. De même, afin de ne pas inclure une variable spécifique à l'entreprise, les résultats provenant de programmes organisés pour une entreprise donnée ont également été supprimés. Notre échantillon final est donc composé uniquement des participants aux programmes de développement du leadership dirigés par le Pr. Kets de Vries à l'INSEAD.

D'autre part, la base étant enregistrée automatiquement, un nettoyage s'est avéré nécessaire pour exclure les questionnaires non terminés, les personnes ayant abandonné avant de valider leur réponses, etc. Ainsi, dans le cas où le questionnaire abandonné correspondait au manager en formation – il est important d'exclure de l'échantillon l'ensemble de ses observateurs – même si ceux-là avaient complété l'évaluation.

Dans la suite de ce travail nous nous référerons au self (selves) pour désigner le manager qui suit le programme et qui a été l'objet de l'observation à 360°. Les Observateurs sont, eux, les personnes ayant accepté de répondre au questionnaire concernant le self.

Ces observateurs sont répartis en 4 catégories : les *Co-Workers*, sont des collaborateurs du self, ils occupent dans l'entreprise des postes hiérarchiquement similaires à celui du self. Les supérieurs (*Superior*) sont des personnes occupant un poste hiérarchiquement supérieur à celui du self. Les *Direct Reports* sont des subordonnés du self, ils occupent donc un poste hiérarchiquement inférieur au self. La dénomination *Direct Report* a été finalement préférée à celle de « *Subordinate* », qui avait été utilisée dans les premières versions du questionnaire. Enfin, une dernière catégorie Autre (*Other*) a été proposée pour désigner tout autre type d'observateurs tels que les clients, les fournisseurs ou les prestataires.

L'association d'un observateur avec une catégorie a été réalisée lors du remplissage du questionnaire, aucune autre option n'était proposée au participant.

A. Les critères d'inclusion dans l'échantillon

Toute personne ayant des non-réponses aux questions obligatoires a été enlevée de l'échantillon (dans le cas où il s'agissait d'un *SELF* tous ses observateurs ont également été enlevés).

Seules les personnes ayant répondu en utilisant la version Anglaise du GELI ont été retenues (si un *OBSERVATEUR* a utilisé une version en français ou en russe il a été enlevé, si un *SELF* a utilisé une version en français ou en russe lui et tous ses observateurs ont été enlevés).

Si du fait de leur participation à plusieurs programmes, un même *SELF* a été évalué plusieurs fois au cours des années 2003-2006, seule la première de ses évaluations a été gardée. Si un *SELF* a également joué le rôle d'observateur pour un autre sujet, son

observation a été enlevée. Ainsi, que ce soit en tant que *SELF* ou *OBSERVATEUR*, un même sujet n'a rempli le questionnaire qu'une seule fois.

Enfin, tous les selves ayant moins de 3 observateurs ont été supprimés. En effet, dans le cadre des analyses multi-niveaux, la variance Intra est estimée de manière moins robuste s'il y a moins de 3 évaluations.

B. Description de l'échantillon

Ainsi, nous avons constitué un échantillon composé de 2441 selves ayant chacun entre 3 et 12 observateurs (moyenne = 7,5, écart type = 1.8)

	Fréquence	Pourcentage
<i>Co-worker</i>	5357	27,4
<i>Direct Report</i>	6804	34,8
<i>Other</i>	2789	14,3
<i>Self</i>	2441	12,5
<i>Superior</i>	2177	11,1
<i>Total</i>	19568	100

Tableau A. 2 Répartition de l'échantillon par type d'observateur.

1. Age

L'âge des répondants varie de 20 à 80 ans avec une moyenne à 41.3 ans et un écart type à 8.5 ans. Sur l'ensemble de l'échantillon 9 personnes ont entre 70 et 80 ans (0.05% de l'échantillon), ce qui peut sembler âgé pour une personne active, mais se comprend si l'on envisage que ces personnes puissent-être les gestionnaires ou les fondateurs de leur propre société.

	Moyenne	Écart-type	Minimum	Maximum
<i>Self</i>	40,35	6,6	24	60
<i>Direct Report</i>	39,1	8,47	20	74
<i>Co-worker</i>	42,07	8,14	21	76
<i>Superior</i>	48,01	7,64	27	72
<i>Other</i>	41,03	8,89	21	80

Tableau A. 3 Description de l'âge de l'échantillon.

Pour les selses l'âge varie de 24 à 60 ans avec une moyenne à 40, ce qui est similaire aux subordonnés (*Direct Report*), aux collaborateurs (*Co-worker*) et aux observateurs de la catégorie autre (*Other*) dont la moyenne d'âge se situe également entre 39 et 42 ans.

Les supérieurs (*Superior*), quant à eux, présentent une moyenne d'âge légèrement plus élevée qui se situe à 48 ans.

Enfin l'âge maximum (80 ans) est atteint par un observateur de la catégorie autre, et l'âge minimum (20 ans) par un subordonné.

3. Genre

Dans l'ensemble les hommes sont plus représentés que les femmes. Pour les selses ce rapport est de 80% d'hommes pour 20% de femmes. Pour les observateurs le rapport est plus équilibré pour les catégories Subordonné et Autre (66% d'hommes pour 33% de femmes), mais il est encore plus extrême pour les observateurs de la catégorie Supérieur (90% d'hommes pour 10% de femmes).

	Non renseigné	Femme		Homme	
	%	N	%	N	%
<i>Self</i>	0,80%	488	20,0%	1933	79,2%
<i>Direct Report</i>	0,20%	2269	33,3%	4523	66,5%
<i>Co-worker</i>	0,40%	1278	23,9%	4056	75,7%
<i>Superior</i>	1,10%	215	9,9%	1937	89,0%
<i>Other</i>	0,20%	924	33,1%	1860	66,7%

Tableau A. 4 Répartition de l'échantillon par genre selon le type d'observateur.

4. Nationalités

Cinq pour cent des participants, n'ont pas renseigné cette information. 145 nationalités sont représentées dans l'échantillon mais 134 d'entre elles englobent chacune moins de 2% des répondants, alors que neuf pays représentent plus de 50% des répondants (en ordre décroissant d'importance dans l'échantillon : Britanniques, Allemands, Néerlandais, Américains, Australien, Danois, Français, Russes et Suédois).

Les 5 pays les plus représentés pour les catégories d'observateur sont le Royaume-Uni, l'Allemagne, les Pays Bas, les États Unis et l'Australie.

Pour les sels, les 5 pays les plus représentés sont le Royaume-Uni, les Pays Bas, l'Allemagne, la France et le Danemark.

	Non renseigné	Nombre de pays représenté	% Pour les 5 pays les plus représentés *
<i>Self</i>	6,40%	90	37% **
<i>Direct Report</i>	6,1%	117	34% *
<i>Co-worker</i>	4,8%	107	38% *
<i>Superior</i>	5,3%	77	44% *
<i>Other</i>	4,6%	96	38% *

Tableau A. 5 Répartition de l'échantillon par nationalité selon le type d'observateur.

* Royaume-Uni, l'Allemagne, les Pays Bas, les États Unis et l'Australie.

** Royaume-Uni, les Pays Bas, l'Allemagne, la France et le Danemark

5. Industries

Concernant le domaine d'activité de l'entreprise, moins de 1% des répondants n'ont pas renseigné la question. Celle-ci était proposée sous forme d'une liste déroulante de 53 activités. Bien que les 53 domaines d'activité soient représentées, 7 domaines concentrent plus de 50% des répondants. En ordre décroissant d'importance dans l'échantillon : banque, alimentaire, hydrocarbures, consulting, pharmaceutique, santé et services sociaux et énergie.

	Non renseigné	Les 3 premiers secteurs	Poids de ces 3 secteurs dans l'échantillon
<i>Self</i>	1,30%	Banque, Alimentaire, Hydrocarbures	31%
<i>Direct Report</i>	0,5%	Banque, Alimentaire, Autres Services	30%
<i>Co-worker</i>	0,8%	Banque, Hydrocarbures, Pharmaceutiques	29%
<i>Superior</i>	2,0%	Banque, Pharmaceutiques, Alimentaire	28%
<i>Other</i>	0,4%	Banque, Consulting, Autres Services	29%

Tableau A. 6 Répartition de l'échantillon par Industries, selon le type d'observateur.

Pour toutes les catégories d'observateurs et pour les selves, les activités de la banque représentent le premier groupe professionnel, dont le poids varie de 14 à 18%.

Dans ce chapitre de méthodologie, nous avons présenté le GELI (*Global Executive Life Inventory*) – outil psychométrique composé de 12 échelles permettant d'évaluer des qualités propres au leadership. Cet instrument présente de bonnes qualités métriques qu'il s'agisse de fidélité ou de validité.

Comme les deux dernières échelles du GELI permettent aux évaluateurs de ne pas répondre à une question s'ils n'ont pas pu observer un comportement donné du manager – seules les 10 premières échelles de l'outil seront utilisées dans cette thèse. Le GELI est complété, d'une part, par le Self (le manager suivant le programme de formation et de développement personnel) et, d'autre part, par un ensemble d'observateurs répartis en quatre catégories : *Direct Report* (les subordonnés), *Superior* (supérieur), *Co-Worker* (collaborateurs) et *Other* (autre). Notre échantillon final est composé de 2441 Selves, évalués chacun par 3 à 12 observateurs. Il s'agit à près de 70% d'hommes, âgés en moyenne de 45 ans et issus d'un pays européen, des États Unis ou d'Australie.

Résultats

Depuis le début des années 2000, de nombreux auteurs ont mis en avant l'importance d'appréhender le leadership en tenant compte de son aspect fondamentalement multi-niveaux (Hunt & Dodge, 2000).

Pourtant, malgré l'accord du milieu scientifique sur l'importance du multi-niveaux dans le leadership, les différentes méthodes associées ont rarement été opérationnalisées (Yammarino, Dionne, Uk, Chun, & Dansereau, 2005). Ainsi, dans leur méta-analyse, Yammarino et ses collaborateurs (2005) notent que seuls 30% des publications reliées au Leadership entre 1995 et 2004 évoquent la question du multi-niveaux, que ce soit au niveau théorique ou empirique.

De même, dans sa revue de littérature plus récente, (Markham, 2010) n'a recensé que 28% des articles empiriques pour la période 2006 – 2010 qui utilisent des techniques multi-niveaux dans leur études des résultats.

Cette approche tenant compte du contexte (également appelés « super-structures » par Hunt et Dodge) est une nécessité conceptuelle pour l'étude du leadership qui, par définition, inclut des managers avec leurs subordonnés, appartenant à des départements et entreprises différentes et dans le contexte de leadership global, évoluant dans des pays et des cultures différentes.

Comme nous l'avons rappelé précédemment, les résultats des questionnaires à 360 degrés, présentent une structure des données particulière où les observateurs sont imbriqués dans les selves. Cela à une conséquence importante pour le choix des méthodes de traitement des données car ces données ne sont pas indépendantes. Or de très nombreuses statistiques utilisées en psychologie posent comme prérequis une indépendance des observations pour obtenir une estimation sans biais.

Les approches hiérarchiques ou multi-niveaux, ont été tout spécialement développées pour analyses des structures de données telles que la nôtre. Dans la première partie de ce chapitre nous détailleront comment cette approche permet d'estimer la variance à chacun des niveaux d'analyses (dans notre cas entre les selves d'une part et entre les observateurs de ces selves d'autre part).

Dans la seconde partie, nous appliquerons ces modèles à nos données et nos analyses viseront à mettre en lumière le pouvoir explicatif de différentes variables évoquées dans le chapitre précédent (Chapitre 4). En effet, dans les études portant sur les analyses à 360°, les variables comme le genre, l'âge ou la catégorie d'observateur ont souvent été évoquées pour expliquer la variabilité des évaluations, nous allons donc vérifier leur effet réel dans le cadre de l'analyse multi-niveaux

Chapitre 5. Les analyses multi-niveaux

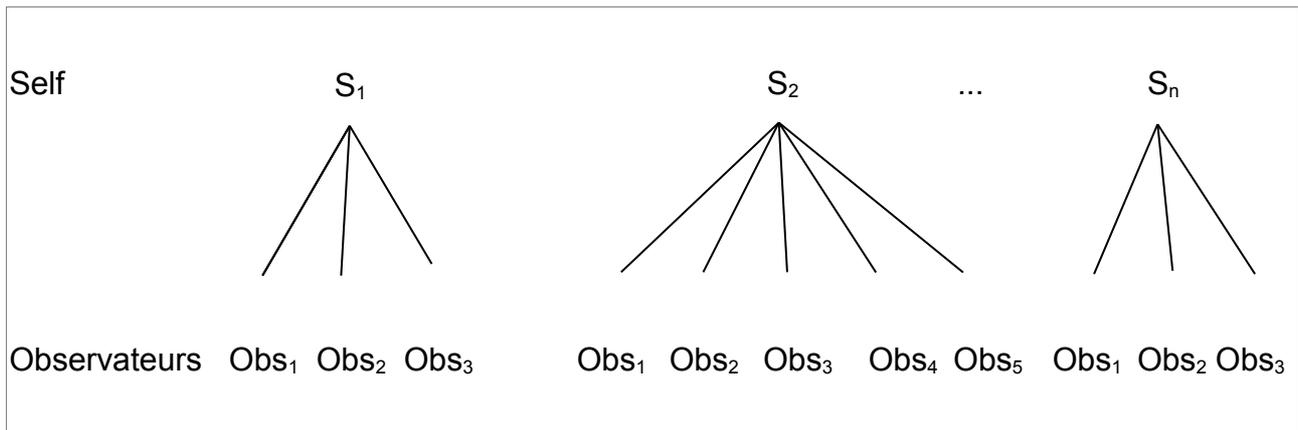
Les analyses hiérarchiques ou analyses multi-niveaux ont été développées et sont largement utilisées dans les domaines des sciences sociales et notamment de l'éducation. Elles permettent de travailler au niveau individuel tout en prenant en compte l'environnement dans lequel l'individu vit. Grâce à ces méthodes, nous pouvons étudier de quelle manière l'environnement socio-économique des individus influe sur les liens statistiques observées au niveau individuel, comme par exemple les liens entre certaines méthodes d'enseignement et les réussites des élèves (Aitkin, Anderson, & Hinde, 1981). Ainsi, en 1981, Aitkin a remis en cause la conclusion d'une étude précédente disant que la méthode d'enseignement « formelle » permettait une meilleure progression des élèves. En effet, dans la première étude, des régressions classiques avaient été utilisées et avaient montré des résultats significatifs. Or, en reprenant les mêmes données, mais en tenant compte des groupes d'appartenance des élèves (classes, professeur) Aitkin a montré que les résultats jusqu'alors significatifs disparaissaient, ce qu'il a appelé un effet « classe ».

1. Les analyses à un et à multiples niveaux

L'analyse multi-niveaux va donc distinguer d'une part les unités au niveau inférieur du système d'analyse, qui sont souvent les individus et qui sont dans notre cas les observateurs. D'autre part, les groupes dans lesquels ces individus sont associés, comme par exemple les classes et dans notre exemple les selves.

Par convention, dans les analyses multi-niveaux, les catégories inférieures (observateurs) seront au niveau 1 et les groupes d'observateurs (par self) au niveau 2.

Ainsi dans le graphique A.1, nous pouvons voir la structure hiérarchique sous-jacente à nos données:



Graphique A.1 Diagramme d'une structure apparié

Il est important de noter qu'une structure hiérarchique telle que la notre, suppose qu'un observateur n'appartienne qu'à un seul self.

A. Les analyses ne tenant pas compte du multi-niveaux

Avant d'aborder le multi-niveaux, nous allons rappeler le modèle statistique de base pour les analyses de régression simple. Pour ce faire nous allons prendre comme exemple les résultats à la dimension *Visioning* comme variable dépendante (y) Équation A.1.

$$y_i = \beta_0 + e_i$$

Équation A. 1 : Modèle d'une analyse de régression simple.

Où y_i est la valeur de *Visioning* pour le i -ème individu ($i=1, \dots, n$), β_0 est la moyenne en *Visioning* dans la population et e_i le résidu du i -ème individu (la différence de cet individu avec la moyenne de la population).

Nous postulons que les résidus sont distribués normalement avec une moyenne à 0 et une variance σ^2 . (Équation A.2)

$$e_i \sim N(0, \sigma^2)$$

Équation A. 2: Résidu

Ainsi, la variance sera d'autant plus grande que les scores individuels y_i vont s'éloigner de la moyenne du groupe β_0 .

B. Les analyses à multiples niveaux

Contrairement à l'exemple précédent, nous allons maintenant ajouter un niveau complémentaire, en introduisant la possibilité de différences dans les moyennes de *Visioning* par groupes. Ainsi, nos individus (observateurs) de niveau 1 sont appariés dans des groupes de niveau 2 (le self qu'ils évaluent). Ainsi, y_{ij} représente la valeur sur la dimension *Visioning* du i -ème individu dans le j -ème groupe.

Dans un tel plan à deux niveaux, nous allons séparer le résidu précédent en deux parties, correspondants aux deux niveau de structure (Équation A.3).

$$y_{ij} = \beta_0 + u_j + e_{ij}$$

Équation A. 3 : Modèle d'une analyse à deux niveaux

Où u_j est le résidu groupe (également appelé effet aléatoire du groupe), e_{ij} le résidu individuel et β_0 est la moyenne générale de *Visioning* tous groupes confondus.

Ainsi, la moyenne pour le groupe j est $\beta_0 + u_j$, ainsi le résidu groupe u_j est la différence entre la moyenne du groupe j en *Visioning* et la moyenne générale β_0 en *Visioning*.

Le résidu du niveau individuel e_{ij} est la différence entre le score en *Visioning* du i -ème observateur avec la valeur moyenne en *Visioning* de son groupe (moyenne sur tous les observateurs d'un même self).

Nous postulons que les résidus aux deux niveaux d'analyses sont distribués normalement avec une moyenne à 0 et une variance σ^2 (Équation A .4).

$$u_j \sim N(0, \sigma_u^2) \quad e_{ij} \sim N(0, \sigma_e^2)$$

Équation A. 4: Résidus du niveau groupe et du niveau individuel.

Ainsi la variance totale est séparée entre deux composantes :

- la variance Inter-groupes (inter-selves) σ_u^2

Correspondant à l'écart entre les moyennes des groupes d'observateurs et la moyenne générale.

- la variance Intra-groupe / Inter-individu (entre les observateurs d'un même self) σ_e^2

Correspondant à l'écart entre les scores des observateurs et la moyenne des scores de leur groupe.

2. Utilisation du multi-niveaux sur les échelles du GELI

Dans l'ensemble des analyses qui vont suivre, nous avons utilisé le logiciel MIWin 2.25 (Rasbash, Charlton, Browne, Healy, & Cameron, 2009). Les paramètres du modèle ont été estimés en utilisant l'algorithme IGLD (*Iterative Generalized Least Square*).

Dans nos modèles multi-niveaux, nous utiliserons les scores aux dimensions du GELI comme variables dépendantes et différentes caractéristique des selves et des observateurs (comme l'âge, le genre, la catégorie d'observateur) comme variables indépendantes.

La majorité de nos variables indépendantes sont catégorielles, or dans les analyses multi-niveaux, lorsqu'on travaille avec des variables catégorielles, l'une des catégories doit être placée en tant que référence et la valeur de son paramètre fixée à 0. L'interprétation de l'intensité et du sens de l'effet des autres modalités se fera donc en référence à cette catégorie.

Si l'on prend l'exemple de la variable Catégories d'observateurs, nous utiliserons la catégorie « supérieur » en tant que catégorie de référence. De ce fait, les valeurs des paramètres pourront être interprétées directement comme les écarts existant entre chacune des autres catégories d'observateur avec la catégorie de référence.

Si l'on s'intéresse à la différence entre les valeurs des paramètres entre des catégories autres que la catégorie de référence, nous devons utiliser le test de Wald. C'est le cas si l'on veut, par exemple, comparer la catégorie Collaborateur avec la catégorie Autre.

Le test de Wald va nous permettre de comparer la différence entre les paramètres, rapportée à l'estimation de l'erreur sur la différence. Ce rapport suit une distribution de Chi^2 , nous permettant de tester l'hypothèse que cette différence n'est pas nulle.

Il est également possible d'introduire des comparaisons plus complexes (entre plus de deux catégories) en utilisant la méthode des contrastes. En fait, dans le cadre d'une comparaison Collaborateur / Autre nous réalisons déjà une comparaison selon la méthode des contrastes. En associant par exemple le contraste -1 à Collaborateur et 1 à Autre,

nous testons ensuite l'hypothèse que cette différence n'est pas nulle.

La valeur de chaque paramètre peut également être testée individuellement en calculant le ratio entre le paramètre et son erreur. Comme nous venons de l'expliquer, ce test n'est qu'un cas particulier du test de Wald (avec 0 coefficients) puisque la valeur de la catégorie de référence est fixée à zéro, et le Chi^2 est le carré d'un Z.

Au niveau global, l'introduction d'un nouveau paramètre affecte la qualité de l'ajustement du modèle aux données. On peut tester cet effet de l'introduction d'un ou de plusieurs paramètres supplémentaires en utilisant, la déviance (-2 Loglikelihood) entre les deux modèles : l'un sans le paramètre d'intérêt et l'autre avec ce paramètre.

La différence entre les déviances suit une distribution de Chi^2 et le nombre de degré de liberté de cette différence correspond à la différence entre le nombre de paramètres du modèle le plus emboîtant avec le modèle contenant le moins de paramètres. Nous pouvons ainsi inférer sur l'importance d'un paramètre donné.

3. Le multi-niveaux est-il nécessaire pour nos données ?

Tout d'abord, nous allons nous intéresser à l'apport du multi-niveaux, à savoir son intérêt dans l'analyse des données de questionnaire à 360°.

Hypothèse :

Au vu de l'intérêt théorique du multi-niveaux pour les données de ce type (Yammarino et al., 2005), nous nous attendons à ce que l'utilisation d'un modèle multi-niveaux permette une réduction significative de la déviance comparativement à un modèle nul (à simple niveau).

Pour ce faire nous allons utiliser comme variables dépendantes les scores aux 10 échelles du GELI, ces scores sont calculés comme la moyenne arithmétique des items les composant. Utiliser une moyenne nous permettra de comparer les résultats sur les différentes échelles dont le nombre d'items varie. Ces scores moyens ont un minimum à un et un maximum à sept.

Exemple pour *Visioning* :

$$Visioning_i = \beta_0 + e_i$$

$$-2 \text{ Loglikelihood} = 43209,951$$

Équation A. 5 Déviance du modèle à 1 niveau pour la dimension *Visioning*.

Pour la dimension *Visioning*, nous pouvons voir qu'en ajustant un modèle ne tenant pas compte de la structure particulière des 360° (structure nichée des observateurs au sein des selves) – nous obtenons une déviance (-2 loglikelihood) de 43209.95 Équation A.5.

$$Visioning_{ij} = \beta_{0j} + e_{ij}$$

$$\text{où } \beta_{0j} = \beta_0 + u_{0j}$$

$$-2 \text{ Loglikelihood} = 41879,482$$

Équation A. 6 Déviance du modèle à 2 niveaux pour la dimension *Visioning*.

Toujours pour *Visioning*, en ajustant un modèle tenant compte de la structure imbriquée des données, nous obtenons une déviance (-2 loglikelihood) de 41879.48 (Équation A.6) .

$$43209,9 - 41879,5 = 1330,4$$

Équation A. 7 Réduction de la déviance entre le modèle à 1 et le modèle à 2 niveaux pour la dimension *Visioning*.

Donc la réduction de la déviance comparativement au modèle nul est de 1330 (Équation 1.7) ce qui, avec un degré de liberté, est significatif à $P < .001$. Cela confirme donc l'importance d'utiliser le multi-niveaux pour appréhender les résultats des questionnaires à 360°.

Dans le Tableau A. 7 nous donnons les résultats pour les 10 dimensions du GELI.

	Modèle de niveau 1	Modèle de niveau 2	Différence
	-2*loglikelihood	-2*loglikelihood	
<i>Visioning</i>	43209,9	41879,5	1330,4 *
<i>Empowering</i>	46054,91	44776,04	1278,87 *
<i>Energizing</i>	45060,3	43339,4	1720,9 *
<i>Designing</i>	46044,5	44805,14	1239,36 *
<i>Rewarding</i>	49023,57	47817,7	1205,87 *
<i>Team Building</i>	46205,23	45055,65	1149,58 *
<i>Outside Orientation</i>	45167,88	44248	919,88 *
<i>Global Mindset</i>	46443,93	44401,29	2042,64 *
<i>Tenacity</i>	43307,94	42169,21	1138,73 *
<i>Emotional Intelligence</i>	48419,31	46716,1	1703,21 *

Tableau A. 7 Déviations (-2Loglikelihood) pour les modèles de niveau un et de niveau deux et les différences correspondantes pour les 10 échelles du GELI.

* différence significative à $p < 0,0001$

Notre hypothèse est donc vérifiée pour les dix dimensions du GELI. En effet, les différences observées dans la déviance entre les modèles à un niveau et les modèles multi-niveaux sont toutes significatives à $P < .001$ avec un degré de liberté. Cela confirme par ailleurs, l'importance d'utiliser le multi-niveaux pour appréhender les résultats de questionnaires à 360°.

4. La variance Inter-selves et Inter-observateurs

Nous allons à présent nous intéresser à la répartition de la variance, à savoir, quelle est la part de la variance totale qui est liée à la variance inter-groupe (i.e. Inter-self) et quelle part de la variance est due aux variations inter-observateurs (i.e. Intra-self).

Hypothèse :

D'après les nombreuses études portant sur l'accord et le désaccord entre les observateurs d'un même self que nous avons étudié dans le chapitre 4, nous nous attendons à ce qu'une part plus importante de la variance soit due aux différences entre les observateurs d'un même self, qu'entre les selves.

Pour ce faire, nous avons réalisé un modèle multi-niveaux pour chacune des dimensions mesurées par le GELI.

Exemple pour *Visioning*

$$Visioning_{ij} = \beta_{0j} + e_{ij}$$

$$\beta_{0j} = 5,481(0,01) + u_{0j}$$

$$u_{0j} \sim N(0, \sigma^2_{u0}) \quad \sigma^2_{u0} = 0,147(0,007)$$

$$e_{ij} \sim N(0, \sigma^2_e) \quad \sigma^2_e = 0,586(0,007)$$

$$-2 \text{ Loglikelihood} = 41879,482$$

Équation A. 8 Variance Inter- et Variance Intra-groupe pris en compte par le modèle pour la dimension *Visioning*.

Pour la dimension *Visioning* (Équation A.8), nous voyons que la variance Inter-groupes (entre les selves) estimée par σ^2_{u0} est égale à 0,15 et la variance Intra-groupe (entre les observateurs de chaque self) estimée par σ^2_e est égale à 0,59.

$$\sigma^2_T = 0,147 + 0,586 = 0,733.$$

Équation A. 9 Variance totale estimée pour la dimension *Visioning*.

La variance totale (σ^2_T) est donc égale à 0.73 (Équation 1.9).

Pour mieux nous rendre compte de la part de variance liée à chacune des sources, nous pouvons calculer le VPC (*Variance Partition Coefficient*) Équation A.10.

$$VPC = \sigma^2_u / (\sigma^2_u + \sigma^2_e)$$

Équation A. 10 *Variance Partition Coefficient*

Ce coefficient varie de 0 (pas de différences inter-groupes $\sigma^2_u = 0$) à 1 (pas de différences au sein des groupes donc variance intra-groupe $\sigma^2_e = 0$).

$$VPC \textit{ Visioning} = \frac{0,147}{0,733} = 0,201$$

Équation A. 11 VPC pour la dimension *Visioning*.

Donc, 20% de la variance est due à la variabilité Inter-groupes et 80% est liée à la variabilité Intra-groupe (Équation A.11).

Dans le Tableau A. 8 nous présentons les résultats pour l'ensemble des 10 échelles étudiées.

	σ^2_u	σ^2_e	% Variance INTER	% Variance INTRA
<i>Visioning</i>	0,15	0,59	20	80
<i>Empowering</i>	0,17	0,7	19	81
<i>Energizing</i>	0,19	0,63	23	77
<i>Designing</i>	0,16	0,7	19	81
<i>Rewarding</i>	0,19	0,83	19	81
<i>Team Building</i>	0,16	0,71	18	82
<i>Outside Orientation</i>	0,13	0,69	16	84
<i>Global Mindset</i>	0,22	0,66	25	75
<i>Tenacity</i>	0,14	0,6	18	82
<i>Emotional Intelligence</i>	0,22	0,77	23	77

Tableau A. 8 Variance et part de variance Intra et Inter pour les 10 échelles du GELI.

Ainsi conformément à notre hypothèse, quelle que soit la dimension, la majeure partie de la variance (dans une proportion 75% / 25%) est liée aux variations intra-groupe, donc entre les observateurs d'un même self.

5. L'importance des variables explicatives

Nous allons à présent étudier les variables explicatives. En effet, maintenant que nous connaissons la part de variance existant aux deux niveaux, nous allons ajouter des variables explicatives et ce, aussi bien au niveau des selves que des observateurs (à chacun des deux niveaux d'analyse), pour tenter d'expliquer au mieux cette variance existante.

Lors de notre revue de la littérature, il est apparu clairement que de nombreuses caractéristiques des selves et / ou des observateurs ont été utilisées pour mieux comprendre les différences observées entre les évaluateurs dans le cadre des questionnaires à 360°. Nous allons donc nous intéresser aux principales : le type d'observateur, l'âge et le genre.

A. La catégorie d'observateurs

Dans les évaluations à 360°, l'un des objectifs est de positionner l'auto-évaluation au regard des évaluations réalisées par les observateurs de chaque catégories. Qui plus est, chacune de ces catégories est supposée avoir des occasions d'observation différentes et donc donner une perspective unique sur les qualités et les défauts d'un manager donné.

Hypothèse :

Au vu de l'importance donnée dans la littérature aux effets des différents types d'observateurs (Hazucha et al., 1982; Hellevik et al., 1992), nous nous attendons à observer des différences significatives dans les évaluations des observateurs appartenant à des catégories différentes évaluant le même self.

- ✧ Nous nous attendons d'une part à une réduction significative de la déviance entre le modèle tenant compte des catégories d'observateur et le modèle nul.*
- ✧ D'autre part, nous nous attendons à des différences quant aux scores attribués par les observateurs des différentes catégories.*

L'analyse qui suit, va introduire dans le modèle les différentes catégories d'observateur : les collaborateurs, les supérieurs, les subordonnés et les autres. Pour ce faire, nous avons créé pour chaque dimension un modèle tenant compte de cette variable, la catégorie

« Supérieur » sera utilisée comme catégorie de référence.

L'exemple pour la dimension *Visioning* est présenté dans l'équation A.12.

$$Visioning_{ij} = \beta_{0j} + \beta_1 Direct\ Report_{ij} + \beta_2 Co - Worker_{ij} + \beta_3 Other_{ij} + e_{ij}$$

$$-2Loglikelihood = 41621,686$$

Équation A. 12 Modèle pour la dimension *Visioning* tenant compte des catégories d'observateurs.

Tout d'abord nous allons nous intéresser à la réduction de la déviance observée entre le modèle nul et le modèle tenant compte des catégories d'observateurs.

$$41879,5 - 41621,7 = 257,8$$

Équation A. 13 Réduction de la déviance entre le modèle nul et le modèle tenant compte de la catégorie d'observateurs pour la dimension *Visioning*.

Pour la dimension *Visioning*, nous avons une réduction de 257.8 (Équation A.13), ce qui est significatif à $P < 0,001$ avec 3 degrés de liberté.

	Différence
<i>Visioning</i>	258 **
<i>Empowering</i>	144 **
<i>Energizing</i>	183 **
<i>Designing</i>	99 **
<i>Rewarding</i>	173 **
<i>Team Building</i>	124 **
<i>Outside Orientation</i>	126 **
<i>Global Mindset</i>	90 **
<i>Tenacity</i>	92 **
<i>Emotional Intelligence</i>	105 **

Tableau A. 9 Réduction de la déviance (-2 Loglikelihood) entre le modèle nul et le modèle tenant compte des catégories d'observateur.

** différence significative à $p < 0,001$ avec 3 degrés de liberté

La réduction de la déviance (-2loglikelihood) entre le modèle nul et le modèle tenant compte de la catégorie d'observateur est significative pour l'ensemble des échelles du GELI. La première partie de notre hypothèse se trouve donc vérifiée.

Nous allons à présent nous intéresser aux valeurs estimées des coefficients de chacune des catégories.

$$Visioning_{ij} = \beta_{0j} + 0,172(0,019)Direct\ Report_{ij} + 0,001(0,02)Co-Worker_{ij} + 0,247(0,023)Other_{ij} + e_{ij}$$

$$\sigma^2_{u0} = 0,145(0,007)$$

$$\sigma^2_e = 0,577(0,007)$$

Équation A. 14 Paramètres et variance inter- et intra-groupe du modèle tenant compte de la catégorie d'observateurs pour la dimension *Visioning*.

Nous pouvons remarquer que certains des coefficients associés aux différentes catégories d'observateur sont significatifs (Équation A.14).

Comme nous l'avons décrit dans la partie 2 de ce chapitre, pour inférer sur la valeur de ces coefficients, on utilise un test qui fait le rapport entre le coefficient et son erreur de mesure (valeur entre parenthèses dans l'équation). Ce rapport suit une distribution de Z, aussi les valeurs supérieures à la valeur repère de la distribution z (i.e. 1.96) sont significatives au seuil de probabilité $\alpha = .05$

Ainsi pour la dimension *Visioning*, les observateurs des catégories Subordonné et Autre, tendent à attribuer des scores de *Visioning* significativement plus élevés que les observateurs de la catégorie Supérieur et Collaborateur (Équation A.15).

$$Subordonnés: \frac{0,127}{0,019} = 9,5 \quad Autres: \frac{0,247}{0,023} = 10,74$$

Équation A. 15 Valeur des paramètres pour la catégorie Subordonné et Autres.

La catégorie Supérieur étant utilisée en catégorie de références, la significativité des coefficients signifie leur différence significative avec ceux de la catégorie d'observateur Supérieur.

	Coefficient Subordonné	Coefficient Collaborateur	Coefficient Autre
<i>Visioning</i>	0,17 ***	0,001	0,25 ***
<i>Empowering</i>	-0,002	-0,15 ***	0,07 **
<i>Energizing</i>	-0,11 ***	-0,15 ***	0,1 ***
<i>Designing</i>	-0,07 **	-0,12 ***	0,07 **
<i>Rewarding</i>	-0,23 ***	-0,18 ***	-0,004
<i>Team Building</i>	-0,03	-0,12 ***	0,12 ***
<i>Outside Orientation</i>	0,02	-0,09 ***	0,13 ***
<i>Global Mindset</i>	0,05 *	-0,003	0,18 ***
<i>Tenacity</i>	0,02	-0,06 **	0,12 ***
<i>Emotional Intelligence</i>	-0,02	-0,08 ***	0,15 ***

Tableau A. 10 Coefficients de chacune des catégories d'observateurs pour les 10 échelles du GELI pour le modèle tenant compte de la catégorie d'observateur.

* significatif à $p < 0,05$

** significatif à $p < 0,01$

*** significatif à $p < 0,001$

En observant ces résultats dans leur ensemble (Tableau A.10), nous pouvons noter que pour toutes les dimensions les observateurs de la catégorie Autre attribuent les scores les plus élevés comparativement aux autres catégories d'observateur.

De même, pour neuf des dix dimensions, les observateurs de la catégorie Autre attribuent des scores significativement plus élevés que les observateurs de la catégorie Supérieur (à l'exception de la dimension *Rewarding and Feedback* où il n'y a pas de différence significative avec la catégorie Supérieur).

La catégorie Autre étant composé d'un groupe assez hétérogène de personnes (fournisseurs, clients...) nous pourrions penser que ceux-ci ne sont pas aussi familiers avec la personne évaluée, ou tout du moins, sont moins fréquemment en contact avec elle. Ils pourraient donc avoir tendance à sur-évaluer le manager ne voulant pas lui nuire.

Les Collaborateurs présentent pour 8 des 10 dimensions des scores significativement inférieurs aux scores attribués par les Supérieurs (pour les dimension *Visioning* et *Global Mindset* les différences ne sont pas significatives).

Enfin, pour les échelles *Energizing, Designing and Aligning* et *Rewarding and Feedback*, les Subordonnés donnent des scores significativement plus faibles que les supérieurs. Il est intéressant de remarquer qu'il s'agit de dimensions qui sont plus importantes pour les subordonnés que pour les supérieurs. En effet, la dimensions Donner de l'énergie et plus encore la dimension de Récompense et de Feedback est plus aisément observable par les subordonnés. Nous pourrions donc croire que les supérieurs auraient tendance à sur-évaluer les qualités de leur managers sur ces dimensions.

Enfin, nous pouvons également remarquer que les supérieurs tendent à donner des scores globalement plus élevés pour un ensemble de dimensions. Cela pourrait être associé au fait que la plupart des managers évalués sont déjà en haut de l'échelle hiérarchique et que leurs supérieurs sont influencés par un effet de prestige (Murphy 1995), ou plus généralement par un effet de Halo.

La seconde partie de notre hypothèse est donc également vérifiée. Dans l'ensemble, les scores des observateurs évaluant le même self sont significativement différents selon leur catégories (collaborateur, supérieur, subordonné ou autre).

Pourtant, malgré l'intérêt de ces interprétations, il est important d'évaluer la part de variance que les différentes catégories d'observateur permettent d'expliquer dans la variance totale des scores aux échelles du GELI.

$$\sigma^2_{T_{categ}} = \sigma^2_{u0} + \sigma^2_e$$

$$\sigma^2_{T_{categ}} = 0,145 + 0,577 = 0,722$$

Équation A. 16 Résidu de variance total après prise en compte de la variables Catégorie d'observateurs.

Pour la dimension *Visioning*, la variance totale $\sigma^2_{T_{categ}}$ après prise en compte de la catégorie d'observateur est égale à 0.72.

$$\frac{(\sigma^2 T - \sigma^2 T_{\text{categ}})}{\sigma^2 T}$$
$$\frac{(0,733 - 0,722)}{0,733} = 0,015$$

Équation A. 17 Part de variance totale d'origine expliquée par l'ajout de la variable catégorie d'observateur.

La part de variance d'origine expliquée par l'ajout de la variable catégorie d'observateur est de 0.015, ce qui correspond à 1,5% de la variance totale d'origine, qui est expliquée par l'ajout de cette variable (Équation A.17).

Dans le Tableau A.11, nous présentons le pourcentage de variance totale d'origine expliquée par la variable Type d'observateurs pour les 10 échelles du GELI.

	Variance totale expliquée par la variable Catégorie d'observateur (en %)
<i>Visioning</i>	1,50%
<i>Empowering</i>	0,81%
<i>Energizing</i>	0,74%
<i>Designing</i>	0,46%
<i>Rewarding</i>	0,68%
<i>Team Building</i>	0,57%
<i>Outside Orientation</i>	0,85%
<i>Global Mindset</i>	0,45%
<i>Tenacity</i>	0,54%
<i>Emotional Intelligence</i>	0,51%

Tableau A. 11 Pourcentage de variance totale expliquée par la variable Catégorie d'observateur pour chacune des 10 échelles du GELI.

En nous basant sur ces résultats, nous pouvons remarquer que bien que l'ajout de cette variable ait permis une réduction significative de la déviance quant à l'ajustement du modèle aux données, la variable catégorie d'observateurs ne permet pas d'expliquer une grande part de la variance et ce, quelle que soit la dimension.

Aussi, sur la base de ces résultats, bien que notre hypothèse concernant les différences ait été vérifiée, il importe de noter que le pouvoir explicatif de ces différences sur la variance totale observée reste très faible.

Il nous apparaît donc d'autant plus important de trouver les variables qui permettent de mieux comprendre la variabilité inter-observateur intra-self mais aussi inter-self.

En adoptant la même démarche, nous allons étudier une autre variable communément utilisée par les chercheurs dans les études sur les 360° : l'âge.

B. Age

Nous allons à présent nous intéresser à l'âge des observateurs et des sels. Comme nous l'avons vu dans la partie descriptive il y a une certaine variabilité dans les âges de notre échantillon, aussi nous allons mesurer dans quelle mesure l'âge des observateurs et du self expliquent des différences dans les scores observés.

Hypothèse :

Nous nous attendons à ce que l'ajout des variables âge du self et âge des observateurs permette une réduction significative de la déviance comparativement au modèle nul.

Également, comme plus les managers évalués sont âgés, plus ils ont d'expérience et plus ils occupent des postes élevés, nous nous attendons à ce que les sels plus âgés obtiennent des scores plus élevés quelle que soit la dimension.

Notre hypothèse est contraire à la théorie des stéréotypes, postulée par Ostroff et ses collaborateurs (Ostroff et al., 2004) qui suppose que les managers plus âgés subiraient l'effet d'un stéréotype négatif dû à leur âge.

Nous n'avons pas d'hypothèses concernant l'effet de l'âge des observateurs.

Enfin, au vu des résultats contradictoires présents dans la littérature, nous nous attendons à ce que cette variable n'aie pas un fort pouvoir explicatif quant à la variance des scores au GELI.

L'exemple pour la dimension *Visioning* est présenté dans l'équation A.18.

$$Visioning_{ij} = \beta_0 + \beta_1 \hat{Age}OBS_{ij} + \beta_2 \hat{Age}SELF_j + e_{ij}$$

$$-2 \text{Loglikelihood} = 40701,434$$

Équation A. 18 Modèle pour la dimension *Visioning* tenant compte de l'âge des observateurs et de l'âge du self.

Nous voyons dans l'équation A.18 que deux caractéristiques ont été ajoutées à notre équation de départ. Il est à noter que la variable associée au niveau d'analyse 2 (celui des observateurs) comporte les indices ij , alors que la variable associée au niveau d'analyse 1 (celui des sels) comporte uniquement l'indice j .

Tout comme précédemment pour la variable Catégorie d'observateur, nous pouvons nous intéresser à la significativité de la réduction de la déviance entre les $-2 \times \log\text{likelihood}$ obtenus pour le modèle nul d'une part et pour le modèle de l'équation A. 18 d'autre part.

$$41879,5 - 40701,4 = 1178,1$$

Équation A. 19 Réduction de la déviance entre le modèle nul et le modèle tenant compte des âges des observateurs et du self pour la dimension *Visioning*.

La différence observée dans la réduction de la déviance est de 1178.1 (Équation A.19) ce qui est significatif à $P < 0,001$ avec deux degrés de liberté. La première partie de notre hypothèse se trouve donc vérifiée.

Nous allons à présent nous intéresser – toujours pour la dimension *Visioning* – aux valeurs des paramètres des variables Âge des observateurs et Âge des selves (Équation A.20).

$$Visioning_{ij} = \beta_{0j} + -0,004 (0,001) \hat{Age}OBS_{ij} + 0,004 (0,002) \hat{Age}SELF_j + e_{ij}$$

$$\sigma^2_{u0} = 0,146 (0,007)$$

$$\sigma^2_e = 0,580 (0,007)$$

Équation A. 20 Paramètres et variance inter- et intra-groupe du modèle tenant compte des âges des observateurs et du self pour la dimension *Visioning*.

$$\hat{Age} observateurs : -\frac{0,004}{0,001} = -4 \quad \hat{Age} Selves : \frac{0,004}{0,002} = 2$$

Équation A. 21 Valeur des paramètres pour les variables Âge des observateurs et Âge des selves.

Les scores attribués par les observateurs en *Visioning* tendent à diminuer significativement quand l'âge des observateurs augmente.

Par contre, les scores attribués aux selves en *Visioning* tendent à augmenter significativement avec l'âge des selves (Équation A.21).

Comme nous pouvons le voir dans le tableau A. 12, ces conclusions ne sont pas identiques pour toutes les dimensions.

	Paramètre Âge Observateur	Paramètre Âge Self
<i>Visioning</i>	-0,004 **	0,004 *
<i>Empowering</i>	-0,001	0,006 **
<i>Energizing</i>	0,001	-0,003
<i>Designing</i>	0,003 **	0,003
<i>Rewarding</i>	0,002 *	0,003
<i>Team Building</i>	0,001	0,003
<i>Outside Orientation</i>	0,001	0,006 **
<i>Global Mindset</i>	-0,003 **	-0,006 **
<i>Tenacity</i>	-0,002 *	0,007 **
<i>Emotional Intelligence</i>	-0,001	0,001

Tableau A. 12 Coefficients des variables âge du self et âge de l'observateur pour les 10 échelles du GELI pour le modèle tenant compte des âges.

* significatif à $p < 0,05$

** significatif à $p < 0,01$

Les selves tendent à obtenir des scores significativement plus élevés, plus ils sont âgés pour les dimensions *Visioning*, *Empowering*, *Outside Orientation* et *Tenacity*. Les autres dimensions à l'exception de *Global Mindset* et de *Energizing* présentent la même tendance mais elle n'est pas significative. Notre hypothèse se trouve donc partiellement vérifiée, pour huit des dix dimensions.

Pour les observateurs, comme pour les scores en *Visioning*, plus les observateurs sont âgés, plus les scores qu'ils attribuent sont significativement faibles pour les dimensions *Global Mindset* et *Tenacity*.

Inversement, plus ils sont âgés, plus les scores qu'ils attribuent sont significativement plus élevés pour les dimensions *Designing and Aligning* et *Rewarding and Feedback*.

Ces résultats sont donc mitigés. Ils ne valident au final ni totalement nos hypothèses, ni complètement la théorie de Ostroff (Ostroff et al., 2004) pour qui les observateurs ont tendance à avoir un stéréotype négatif sur les employés âgés ce qui aboutirait à une sous-évaluation de leurs qualités.

Enfin, bien que significatives, les valeurs de ces coefficients restent très faibles. Il sera intéressant, de se pencher sur ces paramètres dans le modèle final, une fois que nous tiendrons compte de l'ensemble des variables explicatives, pour vérifier leur ampleur.

Nous allons à présent tester la troisième partie de notre hypothèse concernant la part de variance expliquée par les variables âge.

$$\sigma^2_{T_{\text{âge}}} = \sigma^2_{u0} + \sigma^2_e$$

$$\sigma^2_{T_{\text{âge}}} = 0,146 + 0,580 = 0,726$$

Équation A. 22 Résidu de variance total après prise en compte des variables âge des observateurs et âge du self.

L'ajout des variables explicatives âge des selves et âge des observateurs, nous permet de calculer une nouvelle variance totale $\sigma^2_{T_{\text{âge}}}$. (Équation A.22)

$$\frac{(\sigma^2 T - \sigma^2 T_{\text{âge}})}{\sigma^2 T}$$

$$\frac{(0,733 - 0,726)}{0,733} = 0,010$$

Équation A. 23 Part de variance totale d'origine expliquée par l'ajout des variables âge des observateurs et âge du self.

Ces deux variables permettent donc d'expliquer seulement 0,1% de la variance totale d'origine dans les scores de *Visioning* (Équation A.23).

	Différence	Variance totale expliquée par les variables âge (en %)
<i>Visioning</i>	1178 **	0,95%
<i>Empowering</i>	1127 **	0,46%
<i>Energizing</i>	1086 **	0,25%
<i>Designing</i>	1125 **	0,23%
<i>Rewarding</i>	1176 **	0,10%
<i>Team Building</i>	1127 **	0,23%
<i>Outside Orientation</i>	1114 **	0,49%
<i>Global Mindset</i>	1099 **	0,11%
<i>Tenacity</i>	1000 **	-0,04%
<i>Emotional Intelligence</i>	1139 **	0,10%

Tableau A. 12bis Réduction de la déviance (-2 Loglikelihood) entre le modèle nul et le modèle tenant compte des âges des observateurs et des sels, et pourcentage de variance expliqué par les variables âge.

** différence significative à $p < 0,0001$ avec 2 degrés de liberté

Pour l'ensemble des dimensions du GELI, les conclusions sont souvent similaires à celles de *Visioning*. Ainsi, quelle que soit la dimension, l'ajout des variables Âges des sels et Âge des observateurs permettent une réduction significative de la déviance (Tableau A.12bis). Par contre, mis à part les dimension *Visioning*, *Outside Orientation* et *Empowering* (0,5%) la part de variance totale expliquée par ces variables ne dépasse pas 0,3% pour les autres dimensions du GELI.

Notre troisième hypothèse se trouve donc vérifiée : le pouvoir explicatif des variables âge du self et âge des observateur est très faible.

Une autre caractéristique est souvent citée dans la littérature pour expliquer les différences observées dans les évaluations des qualités de leadership : le genre.

C. Genre

Tout comme pour l'âge le genre se retrouve aux deux niveau d'analyse – au niveau des observateurs et à celui des sels. Il est à noter que 0,4% des observateurs et 0,8% des sels n'ont pas renseigné cette information.

Afin de saisir non seulement l'effet du genre de chacun des répondants (self et observateur) mais également l'effet de l'interaction, nous allons construire une nouvelle variable catégorielle croisant les modalités Genre : Homme/Femme et Répondant : Self/Observateur.

Cette nouvelle variable va présenter quatre modalités :

1. Self Homme évalué par Observateur Homme
2. Self Homme évalué par Observatrice Femme
3. Self Femme évaluée par Observateur Homme
4. Self Femme évaluée par Observateur Femme

C'est la modalité Self homme observé par un observateur homme qui sera la catégorie de référence.

Hypothèses :

Nous nous attendons à ce que le genre du manager ainsi que le genre de l'observateur permette une réduction significative de la déviance quelle que soit la dimension étudiée.

- ✧ *Nous nous attendons à ce que le le genre du manager évalué entraîne une différence significative quant aux scores qui lui sont attribués.*

Pourtant au vu des résultats contradictoires des études précédentes, nous ne posons pas d'hypothèse quant au sens de cette différence. En effet, pour une partie des études les femmes seraient favorisées comparativement aux hommes (Bass, Avolio, & Atwater, 1996; Tsui & Gutek, 1984) alors que pour une autre partie des études, ce serait le contraire (Brutus et al., 1993; Petty & Miles, 1976)

- ✧ *Nous nous attendons également à ce que les observateurs présentent un biais de similitude. Ainsi, les femmes observatrices évalueraient les femmes managers plus favorablement que les hommes manager, et les hommes managers évalueraient les hommes managers plus favorablement que les femmes managers.*

L'exemple pour la dimension *Visioning* est présenté dans l'équation A.24.

$$Visioning_{ij} = \beta_{0j} + \beta_1 MaleS - FemaleO_{ij} + \beta_2 FemaleS - MaleO_{ij} + \beta_3 FemaleS - FemaleO_{ij} + e_{ij}$$

$$-2Loglikelihood = 41298,658$$

Équation A. 24 Modèle pour la dimension *Visioning* tenant compte du genre des observateurs et du genre du self.

Nous allons nous intéresser à deux comparaisons. Nous comparerons dans un premier temps l'effet du genre du self pour les observateurs hommes. Pour cela nous pouvons utiliser directement le paramètre β_2 . Dans un second temps, nous comparerons l'effet du genre du self pour les observateurs femmes, nous nous intéresserons alors à la différence entre β_1 et β_3 .

Mais avant de débiter les comparaisons, nous allons nous intéresser à la significativité de la réduction de la déviance due à l'ajout de ces paramètres. Pour ce faire nous allons comparer les réductions dans la déviance (-2 Loglikelihood) entre le modèle nul et ce modèle de l'équation A.24.

$$41879,5 - 41298,7 = 581$$

Équation A. 25 Réduction de la déviance entre le modèle nul et le modèle tenant compte des genres des observateurs et du self pour la dimension *Visioning*.

Nous obtenons une réduction de la déviance de 581 (Équation 1.25) ce qui est significatif à $P < 0,001$ avec 6 degrés de liberté. Ce résultat nous permet de valider la première partie de notre hypothèse.

$$Visioning_{ij} = \beta_{0j} + 0,123(0,017) MaleS - FemaleO_{ij}$$

$$+ -0,122(0,025) FemaleS - MaleO_{ij} + 0,078(0,027) FemaleS - FemaleO_{ij} + e_{ij}$$

$$\sigma^2_{u0} = 0,150(0,007)$$

$$\sigma^2_e = 0,575(0,007)$$

Équation A. 26 Paramètres et variance inter- et intra-groupe du modèle tenant compte des genres des observateurs et du self pour la dimension *Visioning*.

^ Évaluation par les observateurs hommes

Nous pouvons comparer les évaluations faites par les observateurs hommes des managers (selves) hommes et femmes en s'intéressant à la valeur β_2 . La valeur de ce coefficient est de -0.122, ce qui signifie que les observateurs hommes évaluent en moyenne les selves femmes plus négativement que les selves hommes.

$$\beta_2 - \frac{0,122}{0,025} = -4880$$

Équation A. 27 Coefficient β_2

La différence entre les évaluations des selves hommes et femmes, des observateurs hommes est significative à $P < 0,001$ (Équation A.27) ou Test de Wald $\beta_2 = 0$ ($\text{Chi}^2 = 52,8$, 1ddl, $p < 0,001$)

^ Évaluation par les observateurs femmes

Nous pouvons comparer les évaluations faites par les observatrices femmes des managers (selves) femmes et hommes en s'intéressant à la différence entre les paramètres β_1 et β_3 .

$$\beta_1 - \beta_3 = 0,123 - 0,078 = 0,045$$

Équation A. 28 Différence entre les coefficient β_1 et β_3

La valeur de la différence est de 0.045 (Équation A.28) ce qui signifie que les observatrices femmes évaluent plus positivement les managers (selves) hommes que les managers femmes. Le test de l'hypothèse nulle avec la méthode de Wald $\beta_1 - \beta_3 = 0$ est non significatif ($\text{Chi}^2 = 2,289$, 1ddl, NS). Donc cette différence n'est pas significative pour la dimension *Visioning*.

Comme nous pouvons le voir dans le tableau A.13, les femmes observatrices tendent à évaluer les femmes managers plus favorablement que les hommes managers (à l'exception des dimensions *Visioning* et *Empowering*), ces différences cependant ne sont significatives que pour quatre dimensions (*Energizing*, *Designing and Aligning*, *Outside Orientation* et *Tenacity*). Ce résultat confirme, en partie, notre hypothèse concernant l'effet de similitude. Ainsi, bien que l'ensemble des résultats en soient pas significatifs nous observons tout de même une tendance dans ce sens.

Les hommes observateurs, par contre, tendent à évaluer les hommes plus positivement pour trois des dix dimensions seulement (*Visioning*, *Empowering* et *Team Building*) encore que seule la dimension *Visioning* présente un effet significatif. Pour les autres dimensions, bien que le résultat aille dans le sens inverse, aucun de ces effets n'est significatif.

	β_1 Manager Homme évalué par une Femme	β_2 Manager Femme évaluée par un Homme	β_3 Manager Femme évaluée par une Femme	$(\beta_1 - \beta_3)$ Différences entre évaluations d'un manager Homme et Femme par une Femme
<i>Visioning</i>	0,123*	-0,122*	0,078*	0,045
<i>Empowering</i>	0,059*	-0,037	0,028	0,031
<i>Energizing</i>	0,042	0,025	0,110*	-0,068*
<i>Designing</i>	0,044	0,019	0,107*	-0,063*
<i>Rewarding</i>	0,007	0,022	0,056	-0,049
<i>Team Building</i>	0,025	-0,017	0,044	-0,019
<i>Outside Orientation</i>	0,112*	0,040	0,219*	-0,107*
<i>Global Mindset</i>	0,140*	0,018	0,165*	-0,025
<i>Tenacity</i>	0,032	0,012	0,111*	-0,079*
<i>Emotional Intelligence</i>	0,028	0,052	0,084*	-0,056

Tableau A. 13 Coefficients des catégories Homme manager évalué par une femme, femme manager évaluée par un homme, femme manager évaluée par une femme et différence entre les évaluations faites par les femmes observateur pour les managers hommes et femmes.

* significatif à $p < 0.05$

Il semblerait donc que les femmes observatrices tendent à donner des scores plus élevés quel que soit le genre du self comparativement aux hommes observateurs et que l'effet d'interaction ne soit pas si marqué.

Aussi, nous allons réaliser un nouveau modèle tenant compte cette fois-ci, uniquement des deux facteurs principaux Genre de l'observateur et Genre du Self (Équation A.29).

$$Visioning_{ij} = \beta_{0j} + 0,148(0,014) \text{ GenreOBS : Female}_{ij} \\ + -0,086(0,025) \text{ GenreSELF : Female}_j + e_{ij}$$

$$\sigma^2_{u0} = 0,145(0,007)$$

$$\sigma^2_e = 0,580(0,007)$$

Équation A. 29 Modèle pour la dimension *Visioning* tenant compte du genre des observateurs et du genre du self.

Les paramètres des variables (Genre du Self et Genre de l'Observateur) sont significatifs (Équation A.30), cela confirme la première partie de notre hypothèse sur l'effet du genre sur les résultats.

$$\text{Genre observateurs : } \frac{0,148}{0,014} = 10,6 \quad \text{Genre Selves : } -\frac{0,086}{0,025} = -3,4$$

Équation A. 30 Valeur des paramètres pour les variables Genre des observateurs et Genre des selves.

Ainsi les femmes observatrices donnent en moyenne des scores plus élevés sur la dimension *Visioning* que les observateurs hommes, ce qui confirme notre conclusion précédente. Par ailleurs, pour la dimension *Visioning*, les femmes Observées (les selves) tendent à recevoir des scores significativement inférieurs aux hommes observés. Comme nous pouvons le voir dans le tableau A.13 cela est dû à la plus faible évaluation faite par les observateurs Hommes des managers femme (cela n'est vrai que pour trois dimensions)

Pour les autres dimensions, nous pouvons retrouver le résultat dans le tableau A.14.

	Coefficient de Genre- Obs	Coefficient de Genre- Self
<i>Visioning</i>	0,15 **	-0,09 **
<i>Empowering</i>	0,06 **	-0,03
<i>Energizing</i>	0,05 **	0,05 *
<i>Designing</i>	0,06 **	0,03
<i>Rewarding</i>	0,01	0,04
<i>Team Building</i>	0,03 *	0,01
<i>Outside Orientation</i>	0,13 **	0,06 *
<i>Global Mindset</i>	0,14 **	0,04
<i>Tenacity</i>	0,05 **	0,05
<i>Emotional Intelligence</i>	0,02	0,06 *

Tableau A. 14 Coefficients des variables genre du self et genre de l'observateur pour les 10 échelles du GELI pour le modèle tenant compte des genres.

* significatif à $p < 0,05$

** significatif à $p < 0,001$

Nous voyons dans le tableau A.14 que sur l'ensemble des échelles, les femmes observatrices donnent des évaluations plus positives que les hommes observateurs, bien que cet effet ne soit pas significatif pour deux des dimensions : *Rewarding and Feedback* et *Emotional Intelligence*.

D'autre part, les selves femmes tendent à recevoir des scores légèrement plus élevés que les selves hommes sur 8 des 12 échelles (à l'exception de l'échelle *Visioning* et *Empowering*), bien que ces différences ne soient significatives que pour 3 échelles.

Si nous considérons que les dimensions Donner une vision et Donner du pouvoir sont des dimensions sont perçues comme plus masculines alors que l'intelligence émotionnelle, l'orientation vers l'extérieur et le fait de donner de l'énergie plus féminines dans le cadre du management. Alors, l'hypothèse de Brutus et de ses collaborateurs (Brutus, Fleenor, & McCauley, 1999) qui explique que les femmes sont en général évaluées moins positivement que les hommes sur les dimensions considérées comme masculines, serait confirmée.

Cependant, il nous reste à vérifier quelle part de variance les variables genre du self et genre de l'observateur permettent d'expliquer.

Nous pouvons tout comme pour les variables précédentes, calculer la part de variance totale que ces caractéristiques permettent d'expliquer. En calculant d'abord une nouvelle variance totale $\sigma^2_{T_genre}$ (Équation A. 31) :

$$\sigma^2_{T_genre} = \sigma^2_{u0} + \sigma^2_e$$

$$\sigma^2_{T_genre} = 0,145 + 0,580 = 0,725$$

Équation A. 31 Résidu de variance total après prise en compte des variables genre des observateurs et genre du self.

Ensuite, nous pouvons calculer la part de variance totale d'origine expliquée par l'ajout de ces variables.

$$\frac{(\sigma^2 T - \sigma^2 T_{genre})}{\sigma^2 T}$$

$$\frac{(0,733 - 0,725)}{0,733} = 0,011$$

Équation A. 32 Part de variance totale d'origine expliquée par l'ajout des variables genre des observateurs et genre du self.

L'ajout de ces deux variables permet donc d'expliquer seulement 1% de la variance totale dans les scores de *Visioning*. Nous pouvons voir dans le tableau A. 15 que le pouvoir explicatif de ces variables est encore plus faible pour la plupart des autres dimensions du GELI.

	Différence	Variance totale expliquée par les variables genre (en %)
<i>Visioning</i>	581 **	1,09%
<i>Empowering</i>	471 **	0,46%
<i>Energizing</i>	445 **	0,12%
<i>Designing</i>	474 **	1,27%
<i>Rewarding</i>	496 **	0,29%
<i>Team Building</i>	481 **	0,23%
<i>Outside Orientation</i>	545 **	0,98%
<i>Global Mindset</i>	526 **	0,57%
<i>Tenacity</i>	400 **	-0,27%
<i>Emotional Intelligence</i>	479 **	0,30%

Tableau A. 15 Réduction de la déviance (-2 Loglikelihood) entre le modèle nul et le modèle tenant compte du genre des observateurs et de celui des selses, et pourcentage de variance expliqué par les variables genre.

** significatif à $p < 0,001$

Nous pouvons noter que pour l'ensemble des échelles, les variables genre du self et genre des observateurs, permettent une réduction significative de la déviance (-2loglikelihood).

Par ailleurs, pour les dimensions *Visioning*, *Designing and Aligning*, *Outside Orientation* et *Global Mindset*, les caractéristiques de genre des selses et des observateurs permettent d'expliquer une part plus importante de la variance (entre 0,6 et 1,3%) que pour les autres dimensions ($< 0,4\%$). Enfin, il est tout de même à noter que cette part reste faible quelle que soit la dimension.

6. La variance encore libre

Nous allons à présent nous intéresser au modèle complet, tenant compte de l'ensemble des caractéristiques précédemment étudiées.

Ainsi, nous allons prendre en compte l'ensemble des éléments cités dans la littérature : catégorie d'observateur, âge et genre, et ce, aux deux niveaux d'analyse.

Pour la dimension *Visioning*, nous obtenons donc le modèle de l'équation A.33 :

$$\begin{aligned} \text{Visioning}_{ij} = & \beta_{0j} + \beta_1 \text{Direct Report}_{ij} + \beta_2 \text{Co-Worker}_{ij} + \beta_3 \text{Other}_{ij} + \beta_4 \hat{\text{Age}}\text{OBS}_{ij} \\ & + \beta_5 \hat{\text{Age}}\text{SELF}_j + \beta_6 \text{GenderOBS:Female}_{ij} + \beta_7 \text{GenderSELF:Female}_j + e_{ij} \end{aligned}$$

$$-2 \text{Loglikelihood} = 40394,888$$

Équation A. 33 Modèle général pour la dimension *Visioning* tenant compte de la catégorie d'observateur, des âges et des genres des observateurs et du genre du self.

Sans surprise nous pouvons constater une forte réduction de la déviance (-2loglikelihood) entre le modèle nul et le modèle avec les cinq variables explicatives.

$$41879,5 - 40394,9 = 1484,6$$

Équation A. 34 Réduction de la déviance entre le modèle nul et le modèle tenant compte de la catégorie d'observateur, des âges et des genres des observateurs et du genre du self pour la dimension *Visioning*.

Cette réduction de la déviance de 1484 (Équation A.34) est significative à $P < 0,001$ avec 7 degrés de liberté.

$$\begin{aligned} \text{Visioning}_{ij} = & \beta_{0j} + 0,143(0,021) \text{Direct Report}_{ij} + -0,02(0,021) \text{Co-Worker}_{ij} \\ & + 0,215(0,024) \text{Other}_{ij} + -0,00(0,001) \hat{\text{Age}}\text{OBS}_{ij} + 0,002(0,002) \hat{\text{Age}}\text{SELF}_j \\ & + 0,12(0,015) \text{GenderOBS:Female}_{ij} + -0,072(0,025) \text{GenderSELF:Female}_j + e_{ij} \end{aligned}$$

$$\sigma^2_{u0} = 0,144(0,007)$$

$$\sigma^2_e = 0,569(0,007)$$

Équation A. 35 Paramètres et variance inter- et intra-groupe du modèle tenant compte de la catégorie d'observateur, des âges et des genres des observateurs et du genre du self pour la dimension *Visioning*.

Nous pouvons aussi calculer la nouvelle variance totale expliquée par l'ensemble de ces caractéristiques ajoutées au modèle.

$$\sigma^2 T_{\text{model}} = 0,144 + 0,569 = 0,713$$

Équation A. 36 Résidu de variance total après prise en compte des 5 variables du modèle.

Nous sommes donc passés d'une variance totale de 0.733 pour le modèle nul à une variance totale de 0.713 pour ce modèle comportant cinq variables explicatives (Équation A. 36).

$$\frac{(\sigma^2 T - \sigma^2 T_{model})}{\sigma^2 T}$$

$$\frac{(0,733 - 0,713)}{0,733} = 0,027$$

Équation A. 37 Part de variance totale d'origine expliquée par l'ajout des 5 variables explicatives.

Ainsi 2.7% de la variance totale d'origine est expliquée par l'ajout de ces cinq variables (Équation A.37).

	Différence	Variance totale expliquée par les 5 variables (en %)
<i>Visioning</i>	1484,61 **	2,73%
<i>Empowering</i>	1288,96 **	1,27%
<i>Energizing</i>	1289,18 **	1,10%
<i>Designing</i>	1251,31 **	0,93%
<i>Rewarding</i>	1353,87 **	0,88%
<i>Team Building</i>	1262,88 **	0,80%
<i>Outside Orientation</i>	1331,09 **	1,83%
<i>Global Mindset</i>	1254,09 **	0,91%
<i>Tenacity</i>	1109,25 **	0,41%
<i>Emotional Intelligence</i>	1254,93 **	0,61%

Tableau A.16 Réduction de la déviance (-2 Loglikelihood) entre le modèle nul et le modèle tenant compte des catégories d'observateur, du genre et de l'âge des observateurs et du genre et de l'âge des selves, et le pourcentage de variance expliqué par ces 5 variables.

** *significatif à p<0,001*

Quelle que soit la dimension, la réduction de la déviance due à l'ajout de ces cinq variables explicatives est significative à P<0,001 (Tableau A.16).

Pourtant, la part de la variance totale d'origine expliquée par l'ensemble d'entre elles reste assez faible et varie de 0,4% pour la dimension *Tenacity* à 2,7% pour la dimension

Visioning (Tableau A.16). En moyenne, toutes les échelles confondues, la part de variance totale d'origine expliquée est de seulement 1,1%.

Malgré ces valeurs faibles, il serait tout de même intéressant de savoir quelle part de la variance Inter- et Intra-groupe ces caractéristiques permettent de comprendre.

En effet, l'un des grands avantages des analyses multi-niveaux est de séparer dans les calculs les parts de variance Inter (Inter-selves) et Intra (Intra-self, donc entre les observateurs de chaque self).

Nous allons donc chercher dans quelle mesure ces caractéristiques permettent d'expliquer les différences constatées entre les Observateurs évaluant le même self.

Pour la dimension *Visioning* le modèle de l'équation A.33 avec cinq variables explicatives est composé d'une part de variance Inter-groupes (entre les selves) estimée par $\sigma^2_{u0} = 0,144$, et d'une part de variance Intra-groupe (entre les observateurs pour chaque self) estimée par $\sigma^2_e = 0,569$.

$$\frac{(\sigma^2_e - \sigma^2_{eMODEL})}{\sigma^2_e}$$

$$\frac{(0,586 - 0,569)}{0,586} = 0,03$$

Équation A. 38 Part de variance Intra-groupe d'origine expliquée par l'ajout des 5 variables explicatives.

Aussi, comme le montre l'équation A.38, seul 3% de la variance inter-observateur intra-self pour la dimension *Visioning* a pu être expliquée par l'ajout des cinq variables explicative.

	% de la variance Intra-self d'origine expliquée par les 5 variables
<i>Visioning</i>	2,90%
<i>Empowering</i>	1,15%
<i>Energizing</i>	1,59%
<i>Designing</i>	1,14%
<i>Rewarding</i>	1,32%
<i>Team Building</i>	1,12%
<i>Outside Orientation</i>	1,45%
<i>Global Mindset</i>	1,06%
<i>Tenacity</i>	0,33%
<i>Emotional Intelligence</i>	0,65%

Tableau A. 17 Pourcentage de la variance Intra, expliquée par les variables : catégorie d'observateur, âge et genre du self, âge et genre de l'observateur.

Nous voyons ainsi que pour l'ensemble des autres dimensions la part de la variance Inter-observateurs Intra-self expliquée par ces cinq caractéristiques est encore plus faible que pour la dimension *Visioning* et varie entre 0,3 et 1,6% (Tableau A.17)

Bien qu'une partie de la variance ait pu être expliquée par ces caractéristiques, la part restée libre est encore très importante, surtout en ce qui concerne la variance entre les observateurs d'un même self (variance Intra).

Nous allons donc à présent nous intéresser à cette variabilité entre les observateurs d'un même self et plus spécifiquement à la notion d'accord entre les observateurs.

Au vu de la variabilité observée, on peut se demander si ces observateurs sont en accord les uns avec les autres lors de leurs notations ? Mais aussi si cet accord / désaccord est stable selon les dimensions mesurées ? Et enfin s'il y a des caractéristiques particulières, propres aux observateurs qui seraient en désaccord ?

Dans ce premier chapitre d'analyse, nous avons pu montrer l'importance de l'approche multi-niveaux pour appréhender la spécificité des résultats issus des questionnaires à 360°. En effet, quelle que soit la dimension concernée, l'approche tenant compte de la structure imbriquée spécifique à ce type de démarche a permis une réduction significative de la déviance. De même, cette approche multi-niveaux nous a permis de séparer la variance Inter (inter-self) de la variance Intra (Intra-self et donc inter-observateurs).

Dans un second temps, nous nous sommes intéressées aux variables usuellement citées dans la littérature dans le cadre des études portant sur les 360°, pour expliquer les différences observées dans les évaluations. Nous nous sommes intéressés à la catégorie d'observateur, à l'âge (du self et des observateurs) et au genre (du self et des observateurs). Chacune de ses variables a permis une réduction significative de la déviance comparativement au modèle nul – et présente donc un caractère significatif. Nous avons également retrouvé des résultats présents dans la littérature concernant l'effet de la catégorie d'observateur et l'effet du genre. Pourtant, dans la dernière partie de ce chapitre, nous avons pu montrer que malgré la réduction significative de la déviance entre le modèle nul et un modèle tenant compte de l'ensemble des 5 variables que nous venons de citer – la part de variance expliquée par celles-ci reste très faible et se situe entre 1 et 3%. Aussi, comme moins de 3% de la variance Intra-self (entre les observateurs d'un même self) a pu être expliquée grâce à ces 5 variables, nous allons à présent nous intéresser aux écarts entre les évaluations faites par les différents observateurs d'un même self, et plus précisément à la notion d'accord et de désaccord inter-observateur.

Chapitre 6. Étude de la consistance inter-observateurs

Comme il a été rappelé dans la partie théorique, la majorité des études portant sur les 360° se sont peu penchées sur la consistance au sein des observateurs évaluant un même self. Or, au vu des résultats que nous avons obtenus précédemment, nous savons qu'il y a une grande part de variance libre qui n'est pas expliquée par les variables classiquement utilisées. Aussi, nous allons à présent nous interroger sur l'accord et la consistance inter-observateurs intra-self.

Dans le cadre des 360°, chaque Self a un certain nombre d'observateurs qui l'évaluent (dans notre cas ce nombre varie de 3 à 12 observateurs par self). Nous allons donc appliquer les méthodes de la consistance inter-juges à nos données issues de questionnaires à 360°.

1. Les indices statistiques classiques de l'accord et de la consistance inter-juges.

Historiquement, l'intérêt de l'accord inter-juges reposait sur un objectif précis : s'assurer par des moyens statistiques de l'équivalence des juges de sorte à pouvoir, dans l'idéal, en diminuer le nombre tout en garantissant l'exactitude et la stabilité de l'évaluation.

Deux types de paramètres sont couramment étudiés dans la littérature : la fidélité inter-juges et l'accord inter-juges.

Mesurer la fidélité inter-juges consiste à vérifier la consistance relative des notations effectuées par les multiples juges sur de multiples objets. La question posée est donc de savoir si les juges ont ordonné des multiples objets d'une manière consistante avec les autres juges. Pour répondre à cette question les méthodes utilisées sont de type corrélationnelles.

L'indice statistique premier pour estimer la fidélité entre les juges est le coefficient de corrélation Bravais Pearson (Équation B.1). Cet indice permet de mettre à jour une relation linéaire entre deux variables, dans notre cas entre deux juges évaluant le même objet (self).

$$r = \frac{(\text{COV}_{(X,Y)})}{(\sigma_X * \sigma_Y)}$$

Équation B. 1 Corrélation Bravais-Pearson.

Où $\text{COV}_{(X,Y)}$ est la covariance entre X et Y et σ l'écart-type respectivement de X et de Y.

Mesurer l'accord inter-juges consiste à s'intéresser au consensus absolu dans les scores par de multiples juges pour un ou de multiples objets. Dans ce cas, la question posée est donc de savoir si les scores des juges sont interchangeables ou équivalents dans leur valeurs absolues. Pour ce faire, les méthodes statistiques utilisées s'intéressent à la dispersion intra-groupe.

Un des indices classiquement utilisé pour estimer en même temps l'accord et la fidélité est le coefficient de corrélation Intraclasse (ICC – *Intraclass Correlation Coefficient* ; Shrout et Fleiss, 1979).

Les ICC peuvent être interprétés comme la part de la variance observée dans les évaluations qui sont dues à des différences systématiques entre les objets évalués, comparée à la variance totale dans les évaluations.

Quand on souhaite avoir une estimation de la fidélité et de l'accord sur de multiples objets (items) évalués par un ensemble de juges différents (observateurs) dans chacun des groupes et ce sur une échelle d'intervalle (dans notre cas une échelle de type Lickert), nous pouvons calculer les corrélations Intraclasse. Celles-ci correspondent à une analyse de la variance (ANOVA) *one way*, dans le cas où les objets sont traités comme un effet aléatoire (Équation B.2).

Ainsi :

$$ICC(1,1) = \frac{(BMS - WMS)}{(BMS + (k-1)WMS')}$$

Équation B. 2 Corrélation Intraclasse *One Way* ANOVA

Où BMS (*Between-targets Mean Square*) correspond au carré moyen Inter-groupe et WMS (*Within-targets Mean Square*) correspond au carré moyen Intra-groupe et k au nombre de juges jugeant chaque cible.

Un résultat élevé correspondrait à un niveau de consensus entre les juges et à une bonne consistance dans les évaluations (McGraw et Wong, 1996).

Nous pouvons également calculer une corrélation Intraclasse sur les évaluations moyennes par groupe (Équation B.3). En ce qui nous concerne, les observateurs n'ayant pas systématiquement jugé plusieurs selves, cette approche ne peut pas s'appliquer à nos données. Cette démarche permet d'estimer si l'évaluation moyenne donnée par un groupe de juges est fidèle. Cet indice permet donc d'estimer la stabilité des évaluations moyennes de k juges (Shrout et Fleiss, 1979)

Ainsi,

$$ICC(1, k) = \frac{(BMS - WMS)}{(BMS)}$$

Équation B. 3 Corrélation Intraclasse sur des évaluations moyennes.

2. Application aux résultats des questionnaires à 360°

Nous allons dans cette partie nous intéresser à l'accord et à la fidélité de nos observateurs en étudiant d'une part les corrélations entre les évaluations des observateurs et d'autre part les coefficients de corrélation intraclasse.

A. Les corrélations Inter-Observateur

Heidemeier et Moser en 2009, lors de leur méta-analyse de 128 études indépendantes, ont par exemple trouvé une corrélation Bravais-Pearson d'ensemble de .22 entre les évaluations des selves avec leurs supérieurs.

Nous allons à présent étudier les liens entre les évaluations des différents observateurs d'un même self. Pour chacune des 10 échelles du GELI, nous allons calculer la corrélation Bravais-Pearson moyenne, l'écart type et la corrélation médiane, sur l'ensemble des observateurs pris deux par deux.

Ainsi, pour la dimension *Visioning*, nous obtenons les résultats présentés dans le Tableau B.1 :

	Médiane	Moyenne	Écart-type
Question 1	0,2	0,03	0,21
Question 2	0,19	0,05	0,21
Question 3	0,17	0,04	0,17
Question 4	0,17	0,23	0,12
Question 5	0,14	0,04	0,15
Question 6	0,1	0,04	0,11
Question 7	0,13	0,04	0,13
Question 8	0,15	0,04	0,15

Tableau B. 1 Corrélation moyenne, médiane et écart-type pour chaque paire d'observateur pour les 8 items de la dimension *Visioning*.

Nous voyons que pour la question une de l'échelle *Visioning*, en moyenne, les observateurs pris deux par deux ont obtenu une corrélation de .22 entre leurs évaluations du même manager.

Question par question, les résultats varient de .1 à .2. Pour l'ensemble de l'échelle, nous obtenons ainsi pour la dimension *Visioning* les résultats suivants (Tableau B.2):

Corrélation moyenne <i>Visioning</i>	0,16
Écart-type corrélation <i>Visioning</i>	0,03

Tableau B. 2 Corrélation moyenne et écart-type pour la dimension *Visioning*.

La corrélation moyenne pour l'ensemble des items de la dimension *Visioning* est donc de .16, ce qui est faible. Surtout, cela signifie que deux observateurs jugeant le même self sur les items de la dimension *Visioning*, n'ont, en moyenne, que 2,5% de variance commune.

Pour les autres dimensions, les résultats sont similaires (Tableau B.3) :

	Moyenne	Écart Type	Minimum *	Maximum *
<i>Visioning</i>	0,16	0,03	0,1	0,2
<i>Empowering</i>	0,14	0,02	0,1	0,16
<i>Energizing</i>	0,18	0,02	0,15	0,2
<i>Designing</i>	0,14	0,02	0,13	0,18
<i>Rewarding</i>	0,14	0,01	0,12	0,17
<i>Team Building</i>	0,13	0,02	0,1	0,16
<i>Outside Orientation</i>	0,15	0,02	0,12	0,17
<i>Global Mindset</i>	0,2	0,1	0,14	0,43
<i>Tenacity</i>	0,12	0,02	0,09	0,16
<i>Emotional Intelligence</i>	0,16	0,04	0,1	0,21

Tableau B. 3 Corrélation moyenne, minimum, maximum et écart-type pour les 10 dimensions du GELI.

* corrélation minimum et maximum moyennée sur l'ensemble des paires d'observateur, observée sur les items de chacune des échelles.

L'ensemble des corrélations inter-observateur pour les échelles du GELI sont comprises entre .13 et .20 ce qui correspond à une part de variance commune moyenne entre les observateurs de 2 à 3%.

Bien que ces résultats soient faibles, ils sont similaires à ceux obtenus dans la littérature. Mais au delà de la fidélité, ils nous semble important d'étudier également l'accord à proprement parlé. Pour ce faire, nous allons donc calculer les coefficients de corrélation Intraclasse.

B. Les coefficients de corrélation Intraclasse

Dans le contexte des analyses multi-niveaux, les coefficients intraclasse basés sur une ANOVA à effets aléatoires est l'estimation la plus commune aussi bien la fidélité inter-observateurs que de leur consensus.

Comme cet indicateur mesure à la fois la fidélité et le consensus, des valeurs fortes ne peuvent être obtenues que lorsqu'il y a un consensus absolu et, à la fois, une consistance relative entre les juges. Par contre des valeurs faibles peuvent être obtenues lorsqu'il y a soit un faible consensus, soit une faible consistance, soit les deux (LeBreton, Burgess, Kaiser, Atchley, et James, 2003).

Pour chacun des items de l'échelle *Visioning*, nous avons calculé la corrélation Intraclasse entre les observateurs (Tableau B.4).

	ICC (1,1)
Question 1	0,21
Question 2	0,21
Question 3	0,2
Question 4	0,14
Question 5	0,16
Question 6	0,11
Question 7	0,13
Question 8	0,16

Tableau B. 4 Coefficient de corrélation Intraclasse pour chacun des items de la dimension *Visioning*.

Les résultats sur l'ensemble des 80 items étudiés sont similaires à ceux de la dimension *Visioning*. Pour l'ensemble des items, les coefficients varient de .11 à .26. Aussi, afin de nous assurer que toutes les échelles présentent un niveau d'accord comparable, nous avons choisi de présenter les coefficients Intraclasse échelle par échelle. (Tableau B.5)

	ICC (1,1) minimum	ICC (1,1) maximum
<i>Visioning</i>	0,11	0,21
<i>Empowering</i>	0,11	0,18
<i>Energizing</i>	0,15	0,2
<i>Designing</i>	0,13	0,19
<i>Rewarding</i>	0,14	0,18
<i>Team Building</i>	0,11	0,17
<i>Outside Orientation</i>	0,11	0,16
<i>Global Mindset</i>	0,11	0,26
<i>Tenacity</i>	0,11	0,13
<i>Emotional Intelligence</i>	0,11	0,2

Tableau B. 5 Coefficient de Corrélation Intraclasse minimum et maximum sur les items de chacune des 10 échelles du GELI.

Tout comme pour les corrélations Bravais-Pearson réalisées dans la partie A de ce chapitre, nous pouvons remarquer que pour l'ensemble des 10 échelles, les coefficients Intraclasse (ICC 1,1) présentent des intervalles de variations similaires d'une échelle à l'autre. Sur l'ensemble des 80 items, ces coefficients sont faibles, et ne dépassant que rarement .20. Ces valeurs peuvent être interprétées comme un signe d'absence de consensus entre les évaluateurs (observateurs d'un même self) (Nunally & Bernstein, 1994).

Aussi, qu'il s'agisse de la corrélation Bravais-Pearson ou du coefficient de corrélation Intraclasse, nous en arrivons à la conclusion que l'accord, entre les observateurs est faible, voir très faible, lors de l'évaluation d'un même self et ce quelle que soit la dimension étudiée. Pourtant ce résultat est proche de ceux retrouvées dans la littérature sur l'accord dans les 360. Néanmoins, d'autres auteurs n'entendent pas la question de l'accord de cette manière, et proposent une autre façon d'appréhender ces résultats.

3. Une autre hypothèse pour comprendre l'accord entre les évaluateurs.

A. L'hypothèse de la Restriction de la Variance

De nombreuses méthodes statistiques permettent d'estimer l'ampleur de l'accord entre évaluateurs. Pourtant, dans leur travaux, Le Breton et ses collègues (2003) mettent les chercheurs en garde contre l'utilisation d'indices basés sur l'approche corrélationnelle pour estimer l'accord inter-juges, et ce surtout lors d'études de résultats de 360.

Comme nous l'avons énoncé dans la partie théorique (Chapitre 4), l'équipe de LeBreton a cherché à remettre en cause les conclusions concernant le faible accord entre les observateurs intervenant dans le cadre de 360. En effet, pour ces auteurs, les faibles valeurs des coefficients généralement observées dans les études sur les 360 seraient dues non pas à un manque d'accord réel, mais à un artefact statistique.

Ils se basent sur l'hypothèse de restriction de la variance pour expliquer les faibles coefficients généralement observés. Ainsi, pour les auteurs, le faible accord trouvé dans une majorité d'études est lié aux données mêmes. D'après eux, les 360 seraient plus encore que d'autres résultats de tests, sujets à une restriction de la variance.

Cette restriction de la variance interviendrait à plusieurs niveaux : tout d'abord, les personnes évaluées par les 360 tendent à subir – tout le long de leur carrière – une forte sélection « naturelle ». Les 360 étant destinées majoritairement aux managers, ceux-ci ont souvent fait des parcours similaires, ont subi des processus de sélection et de recrutement centrés sur des qualités et des compétences proches, et se retrouvent donc (si tous ces processus de sélection ont bien fonctionné) à présenter assez peu de variabilité.

D'autre part, au fur et à mesure des sessions de formations et de développement personnel que les managers suivent tout le long de leur carrière, ils finissent là aussi, par perdre le peu de variabilité qui leur restait après le processus de sélection. Enfin, comme cela a été montré par dans de nombreuses études, les résultats des 360 tendent à donner, dans l'ensemble, des scores plutôt favorables, avec une tendance (selves et observateurs confondus) à répondre de manière biaisée vers les scores positifs.

Tous ces éléments, ont conduit l'équipe de LeBreton à mettre en garde contre l'utilisation d'indices basés sur une approche corrélacionnelle. En effet, en ayant recours à ces méthodes (corrélacion de Bravais-Pearson, corrélacion Intraclasse) les résultats obtenus pourraient être biaisés par cette restriction de la variance. Plus spécifiquement : le manque d'accord observé (corrélacions faibles à modérées entre les observateurs) serait d'après l'hypothèse de restriction de la variance dues au manque de variabilité réelle dans les données et non à l'existence d'un accord faible.

Aussi, afin de mesurer l'accord inter-observateur sans que celui-ci soit biaisé, l'équipe de LeBreton a proposé un indice – non influencé par la restriction de la variance – le coefficient r_{WG} .

B. Les coefficients r_{WG}

Cette statistique a été originellement proposée par James, Demaree et Wolf (1984, 1993), pour estimer la fidélité inter-correcteurs, intra-groupe. LeBreton et ses collègues ont adapté son utilisation pour mesurer l'accord inter-observateur, intra-self (donc entre tous les observateurs d'un même self).

Cette statistique permet d'estimer l'accord entre les observateurs, en comparant la variance obtenue entre les multiples évaluateurs (observateurs) avec la variance qu'on pourrait obtenir si les réponses de ces observateurs étaient totalement dues au hasard.

La formule de r_{WG} est présentée dans l'équation B.4.

$$r_{wg} = 1 - \left(\frac{S_x^2}{\sigma^2_E} \right)$$

Équation B. 4 Formule de r_{wg}

où S_x^2 correspond à la variance observée entre au moins deux évaluateurs (observateurs) d'un même self. Ainsi, lorsque les évaluateurs sont parfaitement en accord $S_x^2 = 0$.

σ^2_E , correspond quand à lui à la variance qui pourrait être obtenue si les évaluateurs répondaient au hasard. Ainsi, comme S_x^2 / σ^2_E correspond à la proportion de la variance observée liée aux réponses au hasard, r_{WG} peut être interprété comme une « réduction proportionnelle dans la variance d'erreur » (LeBreton et al., 2003, p. 87).

Aussi, r_{WG} étant basé sur la variance intra-sujet (entre les observateurs d'un même self) ce coefficient sera donc insensible au manque de variabilité entre les selves, et ne sera donc pas affecté par la restriction de la variance. Il sera dès lors une estimation vraie du niveau d'accord entre les observateurs.

Cet indice a été étendu aux situations où un même sujet (Self) est évalué par plusieurs évaluateurs (Observateurs) sur $j = 1$ à J items essentiellement parallèles (LeBreton, James, et Lindell, 2009, p. 129).

L'index multi-item $r_{WG(J)}$ est défini dans l'équation B.5 :

$$r_{wg(J)} = \frac{[J(1 - \frac{S^2_{xj}}{\sigma^2_E})]}{[J(1 - \frac{S^2_{xj}}{\sigma^2_E}) + (\frac{S^2_{xj}}{\sigma^2_E})]}$$

Équation B. 5 Formule de $r_{WG(J)}$

où S^2_{xj} est la moyenne des variances observées pour les J items essentiellement parallèles.

4. Calcul des coefficients r_{WG} .

A. La méthode LeBreton

Pour chacune des 10 dimensions du GELI, pour chacun des 2441 groupes d'observateurs (selves), nous avons calculé les coefficients $r_{WG(J)}$.

En utilisant *Statistical Package for Social Sciences* (SPSS) version 17.0 pour Windows, nous avons transposés les données de sorte à pouvoir calculer les coefficients $r_{WG(J)}$. Cette démarche a été réalisée dimension par dimension et item par item. Pour obtenir le coefficient final de chaque groupe d'observateur pour chacune des échelles, une moyenne a été calculée sur l'ensemble des items de l'échelle.

Cette démarche sera illustrée par le calcul des coefficients pour l'échelle Visioning.

Pour chacun des huit items de la dimension et pour les 2441 self de notre échantillon, nous avons calculé la variance (S^2_{i1}) existante entre leurs observateurs dans leurs réponses à l'item numéro un.

Dans un second temps, cette variance S^2_{i1} a été utilisée pour calculer les quatre r_{wg1} s définis par Le Breton (Équation B.6) :

$$r_{wg1 UN} = 1 - \left(\frac{S^2_{i1}}{4} \right)$$

$$r_{wg1 SS} = 1 - \left(\frac{S^2_{i1}}{2,9} \right)$$

$$r_{wg1 MS} = 1 - \left(\frac{S^2_{i1}}{2,14} \right)$$

Équation B. 6 Formules des coefficients $r_{wg1 UN}$, $r_{wg1 SS}$, et $r_{wg1 MS}$

Enfin, les quatre $r_{wg(j)}$ s pour l'ensemble de l'échelle *Visioning* ont été calculés grâce à \bar{S}_{xj}^2 , moyenne des variances de chacun des items, puis en utilisant les formules des $r_{wg(j)}$ s (Équation B.7)

$$r_{wg(j)UN} = \frac{[7(1 - \frac{\bar{S}_{xj}^2}{4})]}{[7(1 - \frac{\bar{S}_{xj}^2}{4}) + (\frac{\bar{S}_{xj}^2}{4})]}$$

$$r_{wg(j)SS} = \frac{[7(1 - \frac{\bar{S}_{xj}^2}{2,9})]}{[7(1 - \frac{\bar{S}_{xj}^2}{2,9}) + (\frac{\bar{S}_{xj}^2}{2,9})]}$$

$$r_{wg(j)MS} = \frac{[7(1 - \frac{\bar{S}_{xj}^2}{2,14})]}{[7(1 - \frac{\bar{S}_{xj}^2}{2,14}) + (\frac{\bar{S}_{xj}^2}{2,14})]}$$

Équation B. 7 Formules des coefficients $r_{wg(j)UN}$, $r_{wg(j)SS}$ et $r_{wg(j)MS}$

	$r_{wg(J)_{UN}}$	$r_{wg(J)_{SS}}$	$r_{wg(J)_{MS}}$
<i>Visioning</i>	0,93	0,86	0,75
<i>Empowering</i>	0,91	0,82	0,67
<i>Energizing</i>	0,92	0,86	0,73
<i>Designing</i>	0,93	0,87	0,75
<i>Rewarding</i>	0,9	0,82	0,67
<i>Team Building</i>	0,91	0,83	0,67
<i>Outside Orientation</i>	0,93	0,88	0,78
<i>Global Mindset</i>	0,92	0,86	0,73
<i>Tenacity</i>	0,94	0,88	0,79
<i>Emotional Intelligence</i>	0,9	0,81	0,66

Tableau B. 6 Coefficients $r_{wg(J)_{UN}}$, $r_{wg(J)_{SS}}$ et $r_{wg(J)_{MS}}$ pour chacune des 10 échelles du GELI.

$r_{wg(J)_{UN}}$ est la moyenne des $r_{wg(J)_{UN}}$ obtenus pour tous les items de l'échelle, les valeurs <0 et >1 ont été remises à 0. $r_{wg(J)_{SS}}$ est la moyenne des $r_{wg(J)_{SS}}$ obtenus pour tous les items de l'échelle, les valeurs <0 et >1 ont été remises à 0. $r_{wg(J)_{MS}}$ est la moyenne des $r_{wg(J)_{MS}}$ obtenus pour tous les items de l'échelle, les valeurs <0 et >1 ont été remises à 0.

Pour l'ensemble des échelles, les coefficients $r_{WG(J)_{UN}}$ présentent des valeurs moyennes supérieures à .90, les $r_{WG(J)_{SS}}$ ont tous des valeurs au dessus de .80 et enfin, les $r_{WG(J)_{MS}}$ ont des valeurs entre .65 et .75 (Tableau B.6).

Ainsi, au vu de ces résultats, nous pouvons conclure à l'existence d'un accord élevé entre les observateurs et ce pour toutes les dimensions du GELI. Ces conclusions étant inverses de celles obtenues suite au calcul des Corrélations Bravais-Pearson et des coefficients de corrélation Intraclasse, nous pouvons donc pencher pour la validation de l'hypothèse de LeBreton et collaborateurs, à savoir que ces faibles coefficients (corrélations BP et ICC) seraient dus à un problème de restriction de la variance dans les données sans qu'il y aurait de problème d'accord entre les observateurs.

Pourtant l'utilisation des coefficients $r_{WG(j)}$ selon l'approche de LeBreton pose, elle aussi, quelques questions.

B. La mise à 0 des valeurs « Hors-Norme »

Dans leur présentation des coefficients r_{WG} les auteurs conseillent de remettre à zéro toutes les valeurs de coefficients sortant des bornes $[0 - 1]$ afin que l'interprétation soit facilitée. Pourtant, sur les données réelles, la quantité de cas concernés est loin d'être négligeable.

En effet, dû à la formule même utilisée pour calculer les coefficients r_{WG} , un grand nombre de scores sortent de ces limites (c'est à dire qu'ils sont supérieurs à un ou inférieurs à zéro). Dans la formule déjà, nous remarquons que la fonction proposée pour calculer r_{WG} a des propriétés mathématiques spécifiques :

- Lorsque la variance observée (S^2_x) se rapproche de l'estimation de la variance aléatoire (σ^2_E) alors la fonction tend vers 0. Donc, plus la variance observée (S^2_x) est proche de la variance obtenue avec une distribution aléatoire (σ^2_E), plus le résultat de r_{WG} sera proche de 0 (donc sera interprété comme une absence d'accord entre les évaluateurs)
- Lorsque la variance observée (S^2_x) tend vers 0, alors la fonction tend vers 1. Donc, plus la variance entre les observateurs (S^2_x) est faible, plus le résultat de r_{WG} sera proche de 1 et donc sera interprété comme un accord élevé entre les évaluateurs.
- La fonction r_{WG} n'est pas définie pour le cas où le dénominateur (σ^2_E) est égal à 0.

En effet, la division par zéro n'étant pas définie, lorsque σ^2_E s'approche de zéro par les valeurs infiniment petites et positives, alors la fonction r_{WG} prend des valeurs très grandes et négatives.

De même, lorsque σ^2_E s'approche de zéro par les valeurs infiniment petites et négatives, alors la fonction prend des valeurs très grandes et positives. Tout cela bien sûr, alors qu'il n'y a que très peu de changement dans la dispersion réelle des données (S^2_x).

Théoriquement, la logique de cette fonction, voudrait que la dispersion aléatoire (σ^2_E) soit systématiquement supérieure à la dispersion observée (S^2_x). Malheureusement, dans la réalité, lorsque nous suivons LeBreton dans la mise en œuvre et le calcul de r_{WG} , cela ne s'avère pas être le cas.

Ainsi, dans un article récent (LeBreton et al., 2009) les auteurs signalent que les valeurs négatives pourraient être source d'information et être, au fond, un signe de divergences systématiques. Il semble donc que la question des valeurs hors-norme ne soit pas tranchée et que ces valeurs puissent au final avoir un intérêt dans l'interprétation, et mériteraient donc d'être gardées.

Pour faciliter les comparaisons inter-études, LeBreton et ses collègues encouragent les auteurs à transcrire les valeurs moyennes des $r_{WG(j)}$ s par dimension étudiée, pour l'ensemble des trois distributions asymétriques (*skewed*) et pour la distribution uniforme. Or, ces moyennes ne donnent que peu d'information si l'on ne remet pas systématiquement à zéro toutes les valeurs en dehors de l'intervalle [0 – 1]. Aussi, comme nous avons décidé de conserver les valeurs hors-normes en vue d'une étude ultérieure, nous avons choisi de présenter le pourcentage de groupes d'observateurs dépassant le seuil de .7.

Nous pouvons ainsi voir sur moyennes présentées dans le tableau B.7 les effets de la non remise à zéro des coefficients hors-norme : l'indice de tendance central perd tout pouvoir explicatif alors que le pourcentage de groupes présentant un résultat compris entre .70 et 1, donne lui une information tout à fait pertinente.

	$r_{wg(J)_{UN}}$		$r_{wg(J)_{SS}}$		$r_{wg(J)_{MS}}$		$r_{wg(J)_{HS}}$	
	Moy	% >.70						
<i>Visioning</i>	0,94	98,4	> 1	92,4	0,54	76,3	0,72	44
<i>Empowering</i>	0,92	96,9	> 1	87,5	0,88	66,5	0	31
<i>Energizing</i>	0,92	97,6	0,84	91	0,82	74,5	> 1	42,4
<i>Designing</i>	0,9	98	0,79	92,1	0,83	77,7	0,5	45,8
<i>Rewarding</i>	0,91	96,1	0,83	86,1	0,77	67,5	0,15	33,2
<i>Team Building</i>	>1	97,2	0,64	87,8	0,9	67	0,38	30,7
<i>Outside Orientation</i>	0,93	97,9	0,93	92,9	0,89	81	> 1	52
<i>Global Mindset</i>	0,93	98,3	0,84	92,4	0,94	70,1	0,14	37,5
<i>Tenacity</i>	0,93	98,1	0,92	92,7	0,79	82,2	0,31	50,1
<i>Emotional Intelligence</i>	0,91	96,4	0,59	85,2	0,53	66,1	0,88	32,7

Tableau B. 7 Coefficient moyen et pourcentage d'observations présentant des coefficients supérieurs à 0,70 pour chacune des échelles du GELI et pour chacun des coefficients $r_{wg(J)_{UN}}$, $r_{wg(J)_{SS}}$, $r_{wg(J)_{MS}}$ et $r_{wg(J)_{HS}}$.

Nous pouvons donc voir dans le tableau B.7, que le taux d'accord est très élevé, pour les coefficients $r_{wg(J)_{UN}}$ il dépasse 96% des groupes, toutes dimensions confondues, pour les coefficients $r_{wg(J)_{SS}}$ il oscille entre 85 et 93%, pour les coefficients $r_{wg(J)_{MS}}$ entre 67 et 80%, alors que pour les coefficients $r_{wg(J)_{HS}}$ le taux d'accord n'est plus que de 30 à 50%.

Malgré ces conclusions encourageantes, nous ne pouvons ignorer la très large fourchette de variation que nous observons selon le coefficient auquel nous nous référons. En effet, selon que l'on regarde le $r_{wg(J)_{UN}}$ ou le $r_{wg(J)_{HS}}$ le pourcentage d'accord passe de 40% de groupes à plus de 95% des groupes.

Nous devons donc sélectionner parmi les quatre types de coefficients $r_{WG(J)}$ celui que nous allons utiliser pour définir si un groupe d'observateur est ou non en accord. En d'autres mots nous devons choisir la distribution qui sera utilisée en tant qu'estimation de σ^2_E dans le calcul du $r_{WG(J)}$.

C. Le choix de la distribution de référence pour estimer σ^2_E

La statistique proposée par LeBreton et ses collaborateurs (2003), r_{WG} permet de comparer la dispersion observée entre les réponses des observateurs (S^2_x) avec la dispersion qui aurait été obtenue si ces observateurs avaient répondu au hasard (σ^2_E). Pour estimer ces réponses au hasard justement, LeBreton et ses collaborateurs (2009) proposent d'utiliser plusieurs distributions. Ainsi, dans leur article de 2009, les auteurs fournissent une table présentant les valeurs de σ^2_E en fonction du nombre d'ancrages présentes dans les modalités de réponses (les valeurs sont disponibles pour les échelles de réponses allant de cinq à 11 ancrages).

Pour chacune de ces modalités de réponses, les auteurs proposent donc six types de courbes de références. En effet, pour estimer la dispersion due au hasard (σ^2_E), différentes approches pourraient être envisagées.

Si nous prenons l'exemple d'une échelle de réponse en sept points (sept ancrages), pour calculer r_{WG} , nous devons, pour chaque item de la dimension, comparer la dispersion des réponses des observateurs (S^2_x) avec une dispersion aléatoire des réponses (σ^2_E). Cette dispersion aléatoire peut être conçue différemment : nous pourrions envisager d'utiliser comme modèle de réponse aléatoire, une distribution uniforme – cela signifierait que l'on considère que « si les observateurs répondaient au hasard, nous obtiendrions la même probabilité d'utilisation de chacun des sept ancrages de l'échelle de réponse ».

Nous pourrions aussi considérer qu'en répondant au hasard, les observateurs auraient tendance à répondre selon la loi normale – il y aurait donc une plus grande probabilité qu'ils attribuent une note moyenne qu'une note extrême.

Enfin, en suivant la même logique et au vu des connaissances qui sont les nôtres sur le 360 et les biais de réponse existants, nous pourrions estimer qu'en répondant au hasard, les observateurs auraient tendance plus fréquemment à attribuer des scores forts que

faibles (ils utiliseraient plus facilement les ancrages élevés), leur réponses suivraient donc une distribution négativement asymétrique.

Distribution	1	2	3	4	5	6	7	σ^2_E
<i>Slight skew</i> (légèrement asymétrique)	.05	.05	.10	.15	.15	.30	.20	2,90
<i>Moderate Skew</i> (moyennement asymétrique)	.00	.05	.10	.10	.15	.35	.25	2,14
<i>Heavy Skew</i> (fortement asymétrique)	.00	.00	.05	.10	.15	.30	.40	1,39
Normale	.02	.08	.20	.40	.20	.08	.02	1,40
Uniforme	.14	.14	.14	.14	.14	.14	.14	4,00

Tableau B. 8 Variance d'erreur attendue (σ^2_E) et proportion d'individus choisissant chacun des ancrages de réponse.

Données pour les modalités de réponse à 7 ancrages issu de Le Breton J.,M., Senter J., L. (2008) Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*. 11, p. 832.

Nous pouvons remarquer dans le tableau B.7 que la distribution uniforme est celle qui permet d'obtenir la valeur de dispersion aléatoire (σ^2_E) la plus élevée, suivie par la distribution *Slight skew* (légèrement asymétrique) et la *Moderate skew* (moyennement asymétrique).

Pourtant, afin d'utiliser la statistique $r_{WG(J)}$ dans son sens originel, et en respectant son objectif original, il importe de comparer la dispersion obtenue (S^2_x) avec une dispersion aléatoire présentant les mêmes biais. De ce fait, il nous semble particulièrement difficile de retenir la distribution uniforme pour représenter les réponses au hasard des observateurs évaluant des *top managers* dans le cadre d'une approche par 360.

D'après les résultats des recherches précédentes, nous savons que les observateurs prenant part aux évaluations à 360° tendent à faire preuve de biais. Ces biais, en général, ont pour conséquence de faire tendre leurs évaluations vers les scores élevés. Nous pouvons donc dire que dans l'ensemble, les réponses des observateurs tendent à avoir une distribution asymétrique négativement, c'est à dire, déplacée vers la droite.

Dans leurs travaux, LeBreton et son équipe, proposent trois distributions qui tiennent compte de cette tendance. Chacune des distributions intègre un biais plus ou moins sévère dans les réponses (de légèrement asymétrique à fortement asymétrique en passant par moyennement asymétrique).

Comme nous l'avons signalé précédemment, le choix de la distribution de références (distribution qui sera utilisée pour estimer la dispersion due au hasard - σ^2_E) est importante au delà des considérations simplement théoriques. En effet, en fonction de la distorsion (*skewness* / asymétrie) de la distribution choisie (légère, moyenne ou forte), la dispersion associée sera différente (plus ou moins élevée selon le degré d'asymétrie comme cela peut être lu dans le tableau B. 9), et donc, mathématiquement, le nombre de groupes d'observateurs catégorisés comme « en Accord » (ayant un coefficient $r_{wg} > .70$) sera, lui aussi, différent.

Ainsi, pour choisir la distribution à utiliser deux approches nous semblent possibles soit, nous utilisons la même distribution type (*slight skew*, *moderate skew* ou *heavy skew*) pour calculer σ^2_E et donc estimer le hasard pour l'ensemble des 10 dimensions, soit, nous considérons que des biais de différente intensité peuvent affecter les dimensions mesurées, nous utilisons donc, pour calculer σ^2_E et estimer le hasard, des distributions type différentes selon les dimensions étudiées.

	1	2	3	4	5	6	7
<i>Visioning</i>	0%	2%	5%	11%	24%	37%	20%
<i>Empowering</i>	1%	3%	6%	14%	27%	33%	15%
<i>Energizing</i>	0,5%	2%	5%	10%	23%	36%	24%
<i>Designing</i>	0%	2%	5%	15%	28%	35%	14%
<i>Rewarding</i>	1%	3%	6%	16%	26%	32%	15%
<i>Team Building</i>	1%	3%	6%	15%	26%	33%	15%
<i>Outside Orientation</i>	0%	2%	4%	12%	24%	38%	21%
<i>Global Mindset</i>	1%	2%	4%	16%	22%	33%	22%
<i>Tenacity</i>	0%	1%	3%	7%	19%	41%	28%
<i>Emotional Intelligence</i>	1%	4%	6%	14%	25%	33%	16%

Tableau B. 9 Distribution de la répartition des réponses dimension par dimension, pour les 7 ancrages disponibles pour les 17127 Observateurs.

Bien que des différences puissent apparaître pour certaines échelles concernant la répartition des réponses pour les sept ancrages de l'échelle de Lickert (Tableau B.9) – par exemple la dimension *Tenacity* est plus asymétrique que les autres. Dans l'ensemble, nous pouvons observer une tendance similaire dans les réponses quelle que soit la dimension. Cette tendance confirme les biais habituellement mentionnés dans la littérature.

La distribution utilisée pour calculer σ^2_E est en quelque sorte, la distribution qui aurait été obtenue si l'ensemble des observateurs possible (le niveau parent) avaient répondu de manière aléatoire.

Aussi, s'agissant d'une estimation de la répartition des réponses au niveau parent, l'utilisation d'une même distribution pour toutes les échelles du GELI pour calculer σ^2_E et estimer le hasard, nous semble préférable. En effet, les recherches actuelles ne nous permettent pas de postuler à priori des tendances notablement différentes dans les biais de réponse selon les dimensions mesurées, bien que des différences puissent être observées au niveau de notre échantillon (Tableau B.10).

	Moyenne	Médiane	Skewness	Kurtosis
<i>Slight Skew</i> SS	5,00	5,50	-0,74	-0,33
<i>Moderate Skew</i> MS	5,40	6,00	-0,82	-0,32
<i>Heavy Skew</i> HS	5,90	6,00	-0,92	-0,09
<i>Visioning</i>	5,46	6,00	-1,01	1,1
<i>Empowering</i>	5,23	5,00	-0,8	0,52
<i>Energizing</i>	5,53	6,00	-1,04	1,09
<i>Designing</i>	5,28	5,00	-0,84	0,88
<i>Rewarding</i>	5,18	5,00	-0,76	0,32
<i>Team Building</i>	5,24	5,00	-0,81	0,52
<i>Outside Orientation</i>	5,56	6,00	-0,83	0,59
<i>Global Mindset</i>	5,43	6,00	-0,85	0,63
<i>Tenacity</i>	5,80	6,00	-1,02	1,01
<i>Emotional Intelligence</i>	5,21	5,00	-0,79	0,27

Tableau B. 10 Moyennes, médianes, skewness et kurtosis pour correspondant aux biais proposés par LeBreton et ceux observés pour notre échantillon pour les 10 dimensions.

Nous avons choisi d'utiliser la distribution Moyennement Asymétrique (MS – *Moderate Skew*) d'une part parce que cette distribution correspond aux biais décrits dans la littérature, mais également parce que c'est la distribution qui s'approche le plus de nos données.

	Moyenne	Médiane	Skewness
<i>Slight Skew</i> SS	5,00	5,50	-0,74
<i>Moderate Skew</i> MS	5,40	6,00	-0,82
<i>Heavy Skew</i> HS	5,90	6,00	-0,92
Données Total	5,35	6,00	-0,91

Tableau B. 11 Moyenne, médiane et skewness attendus pour chacune des trois courbes asymétriques (légère, moyenne et forte) et pour pour la distribution des données expérimentales.

Si nous observons la distribution moyenne des réponses de l'ensemble des 17127 observateurs sur l'ensemble des 80 questions (correspondant aux 10 dimensions étudiées), la distribution obtenue est proche de la distribution Moyennement Asymétrique (Tableau B.11).

5. Catégorisation des groupes d'observateurs en tant que « Accord », « Désaccord » ou « Hors-Norme »

A. Résultat des analyses

Afin de répondre à notre question initiale, à savoir l'identification des groupes d'observateurs en accord et des groupes d'observateur en désaccord, nous allons catégoriser ceux-ci en fonction de la valeur du coefficient $r_{WG(J)_MS}$.

Nous allons suivre les recommandations de LeBreton (2003) en ce qui concerne les valeurs seuil. Ainsi, les groupes présentant un indice $r_{WG(J)_MS} > 0.70$ seront catégorisés comme « En Accord », d'autre part, les groupes présentant un indice $r_{WG(J)_MS} < 0.70$ seront, eux catégorisés comme « En Désaccord ». Enfin, les groupes présentant des indices $r_{WG(J)_MS}$ inférieurs à 0 ou supérieurs à 1 seront, eux, catégorisés comme « Hors-norme ».

Pour ce qui est du groupe « Hors-Norme », comme nous l'avons présenté dans la partie B de ce chapitre (paragraphe 3), les valeurs les plus extrêmes de r_{WG} correspondent, en fait, à des réalités statistiques proches. En effet, c'est au moment où la fonction r_{WG} s'approche du point où elle n'est pas définie (le cas où le dénominateur est égal à 0) que le résultat va tendre soit vers plus l'infini soit vers moins l'infini.

Ainsi, bien que dans la réalité la variance (S^2_x) de deux groupes soit proche, le fait que celle-ci soit juste supérieure ou juste inférieure à l'estimation de la variance d'erreur σ^2_E , va entraîner des scores $r_{WG(J)_MS}$ très élevés et positifs ou très élevés et négatifs. Ce qui explique le regroupement de ces groupes d'observateurs dans une même catégorie.

Dans le tableau B. 12, nous présentons le pourcentages de groupes d'observateurs dans chacune de ces catégories, pour chacune des 10 dimensions étudiées.

	Pourcentage Accord	Pourcentage Désaccord	Pourcentage Hors-norme
<i>Visioning</i>	76,3	15,4	8,3
<i>Empowering</i>	66,5	19,8	13,7
<i>Energizing</i>	74,4	15,5	10,2
<i>Designing</i>	77,7	13,9	8,4
<i>Rewarding</i>	67,3	18,2	14,5
<i>Team Building</i>	66,4	20,1	13,4
<i>Outside Orientation</i>	80,4	11	7,6
<i>Global Mindset</i>	70,1	16,2	8,6
<i>Tenacity</i>	82,3	10,2	7,5
<i>Emotional Intelligence</i>	65,8	18,4	15,9

Tableau B. 12 Pourcentages de groupes d'observateurs en Accord, en Désaccord ou Hors-Norme, pour chacune des 10 dimensions étudiées.

Pour chacune des dimensions, une large partie des groupes est catégorisée comme en Accord (entre 65 et 80% des groupes), entre 10 et 20% des groupes sont catégorisés comme Désaccord et enfin, le groupe Hors-Norme est composé de 9 à 15% des groupes d'observateurs.

Dans la suite de ce chapitre nous allons nous intéresser à ce qui, mis à part les coefficients de LeBreton, différencie ces groupes.

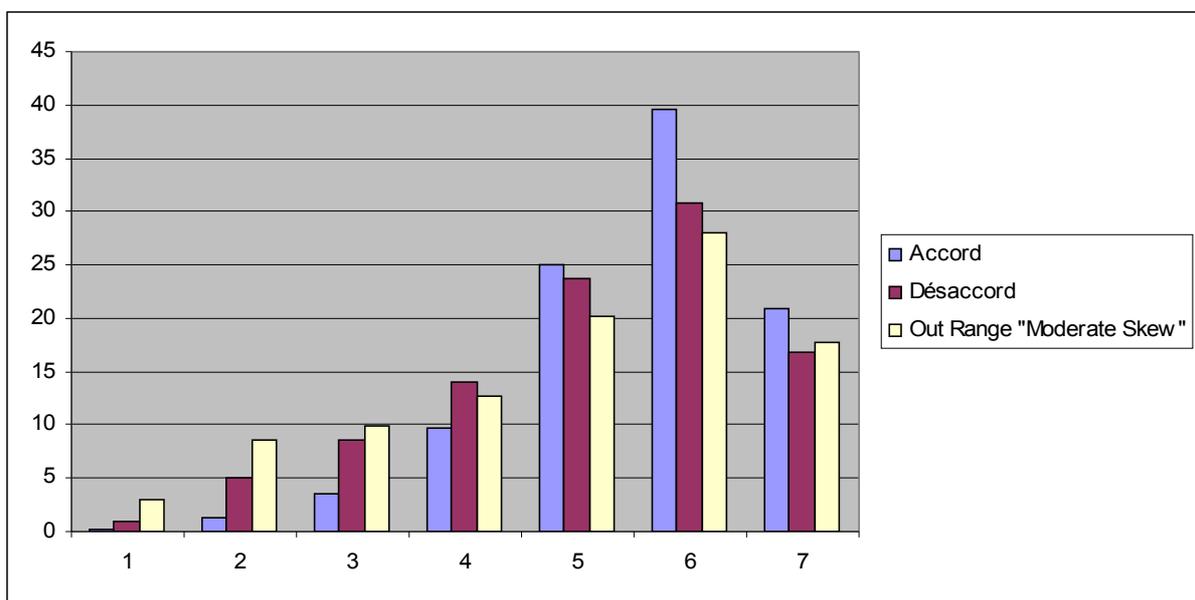
En effet, ces groupes ont été construits sur la base de la variance qui existait entre les observateurs d'un même self. Nous savons que ces groupes présentent plus ou moins de variabilité dans leur réponses, mais comment cette variabilité se traduit-elle en ce qui concerne l'utilisation de l'échelle ?

Dans les groupes en accord, trouve-t-on aussi bien des groupes d'observateurs d'accord entre eux sur la faible performance d'un manager que de groupes d'accord entre eux sur la très bonne performance du self ? Si c'est le cas, nous ne devrions pas observer de différences notables quant à l'utilisation moyenne des sept ancrages de l'échelle de Lickert utilisée pour répondre au questionnaire.

Cependant, nous faisons l'hypothèse que les groupes en désaccord sont, en fait, les seuls groupes ayant en leur sein des observateurs utilisant des ancrages faibles – ce qui crée par ailleurs le désaccord. De ce fait, nous nous attendons à trouver des différences quant à l'utilisation de l'échelle entre les trois groupes d'accord.

B. Utilisation de l'échelle

En prenant l'exemple de la dimension *Visioning*, nous avons calculé les pourcentages d'utilisation de chacun des sept ancrages par les trois groupes (Graphique B.1).



Graphique B. 1 Pourcentage d'utilisation des 7 modalités de réponses aux 8 questions de la dimension *Visioning* par les observateurs, selon leur groupe « d'accord »

Nous pouvons donc remarquer que les observateurs en désaccord se distinguent des groupes en accord par leur utilisation des ancrages de réponse un, deux et trois ; et par leur moindre utilisation de l'ancrage six (Tableau B.13).

	Modalités de réponses avec écarts (Accord vs. Désaccord / Hors-norme)	
	Désaccord > Accord	Accord > Désaccord
<i>Visioning</i>	2 ; 3	6
<i>Empowering</i>	2 ; 3	5 ; 6
<i>Energizing</i>	2 ; 3 ; 4	6 ; 7
<i>Designing</i>	2 ; 3	5 ; 6
<i>Rewarding</i>	3	5 ; 6
<i>Team Building</i>	3	5 ; 6
<i>Outside Orientation</i>	2 ; 3 ; 4	6
<i>Global Mindset</i>	2 ; 3	6
<i>Tenacity</i>	2 ; 3	6
<i>Emotional Intelligence</i>	2 ; 3	5 ; 6

Tableau B. 13 Modalités des réponses présentant le plus de différences quant à leur fréquence d'utilisation, entre les groupes d'observateurs en Accord et ceux en Désaccord.

Aussi, sur l'ensemble des dimensions étudiées, nous observons qu'en moyenne, la fréquence d'utilisation de chacun des sept ancrages (1= « *Not at all* » à 7= « *Very well* ») diffère selon que les observateurs appartiennent au groupe « Accord », « Désaccord » ou « Hors-Norme » et ce, particulièrement pour les modalités de réponse deux, trois et six.

Notre hypothèse se trouve donc vérifiée : les observateurs en accord tendent à être d'accord sur les bonnes compétences des selses uniquement, et c'est la présence d'individus utilisant les ancrages faibles, qui crée le désaccord.

Ces analyses mettent en lumière les différences existant entre les groupes d'observateurs dans leurs évaluations. Pourtant, à cette étape une nouvelle question surgit : les observateurs en accord, le sont-ils toujours ? Les observateurs en accord le sont-ils quelle que soit la dimension mesurée ? Y a t-il une stabilité dans l'accord des observateurs ? Qu'en est-il au niveau général, existe-t-il des groupes toujours d'accord et des groupes toujours en désaccord, ou bien cela est-il variable selon les dimensions ?

6. Généralisation de la consistance entre observateurs

A. Calcul d'un indicateur général $r_{WG(J)}$ Total

Les mêmes observateurs ont évalués le même self sur 10 dimensions, nous pouvons donc légitimement nous poser la question de la stabilité de l'accord entre évaluateurs. En effet, nous avons jusqu'à présent abordé la question de l'accord inter-observateur uniquement de manière « dimensionnelle », c'est à dire échelle par échelle. Il peut donc être intéressant d'aborder la question de l'accord de manière « Globale ».

Pour ce faire, nous allons nous définir des groupes d'accord / désaccord en recalculant un coefficient $r_{WG(J)}$ pour l'ensemble des 10 échelles. Ensuite, nous comparerons les classements obtenus au niveau de chaque échelle avec celui obtenu par la méthode globale.

Bien sûr, il est à noter que l'ensemble des difficultés évoquées précédemment pour les $r_{WG(J)}$, s'appliquent aussi au calcul de ce coefficient général.

Toujours en utilisant la distribution Moyennement Asymétrique pour estimer σ^2_E , nous obtenons la répartition suivante entre les groupes Accord, Désaccord et Hors-Norme (Tableau B14).

	N	Pourcentage
Désaccord	469	19,2
Accord	1570	64,3
Hors-Norme	402	16,5
Total	2441	100

Tableau B. 14 Répartition des 2441 groupes d'observateurs selon leur propension d'accord.

B. Comparaison du groupement « Général » avec le groupement « Par échelles »

Pour effectuer cette comparaison, 10 tableaux de contingence ont été réalisés, comparant le classement des 2441 selves (en fait leur groupe d'observateurs) en trois catégories « Accord », « Désaccord » et « Hors-norme » selon l'approche dimensionnelle d'une part, et générale de l'autre.

Dans le tableau de synthèse B. 15, les résultats pour les 10 échelles sont présentés. La colonne « %1-1 » correspond au pourcentage de groupes (sur 2441) qui ont été catégorisées en Accord par les deux approches (dimensionnelle et générale). La colonne « %Similaire » correspond au pourcentage de groupes d'observateurs classés dans la même catégorie par les deux approches (pourcentage d'accord entre les deux classifications). Enfin, le Chi² et le Tau-b sont présentés pour chaque dimension.

	% 1-1 Total - Dimension	% Similaire 1 ; 0 ; 9	Chi ²	Tau-b
Visioning	57,6%	68,1%	545,33	0,123
Empowering	53,2%	67,4%	672,62	0,162
Energizing	58,2%	69,5%	5161,97	0,120
Designing	58,3%	68,9%	548,59	0,159
Rewarding	54,9%	69,3%	6486,11	0,172
Team Building	55,3%	70,7%	7297,25	0,195
Outside Orientation	58,7%	66,1%	335,06	0,066
Global Mindset	54,6%	63,9%	338,81	0,091
Tenacity	59,1%	65,9%	2272,72	0,097
Emotional Intelligence	54,5%	70,5%	7262,11	0,221

Tableau B. 15 Pourcentage de groupes ayant des Observateurs en catégorisés comme en « Accord » par les deux approches, pourcentage de groupes d'observateurs classés dans la même catégorie par les deux approches, Chi² et le Tau-b pour chaque dimension.

En rouge significatif à $p < 0.0001$

Dans l'ensemble, nous observons un taux d'accord dans le classement (Accord, Désaccord et Hors-norme) qui se situe entre 65 et 75%. Ce qui peut sembler élevé, mais qui indique en même temps l'existence de différences dans l'accord.

Ainsi, des groupes d'observateurs catégorisés comme en Accord pour une des dimension, sont catégorisés comme en Désaccord ou Hors-Norme si l'on prend en compte l'ensemble des dimensions, et ce, dans 25 à 35% des cas.

Pour confirmer notre conclusion, nous allons étudier l'accord en calculant les Tau de Kendall.

C. Stabilité du classement des groupes d'observateurs dans les catégories d'accord, pour les 10 dimensions du GELI.

Comme nous l'avons vu, au niveau de chaque échelle, chaque groupe d'observateur a été catégorisé dans l'une des trois catégories d'accord (i.e. Accord, Désaccord, Hors-norme). Nous allons donc comparer les classements des groupes d'observateur pour les 10 échelles.

		VIS	EMP	ENG	DES	REW	TEAM	OUT	GLB	TEN
<i>Accord Empowering</i>	Corr Coeff Sig.	,094** ,000								
<i>Accord Energizing</i>	Corr Coeff Sig.	,170** ,000	,080** ,000							
<i>Accord Designing</i>	Corr Coeff Sig.	,108** ,000	,085** ,000	,104** ,000						
<i>Accord Rewarding</i>	Corr Coeff Sig.	,061** ,001	,105** ,000	,100** ,000	,122** ,000					
<i>Accord Teambuilding</i>	Corr Coeff Sig.	,117** ,000	,166** ,000	,131** ,000	,128** ,000	,137** ,000				
<i>Accord Out. Orientation</i>	Corr Coeff Sig.	,077** ,000	,059** ,002	,020 ,293	,012 ,531	,024 ,208	,056** ,003			
<i>Accord Global Mind.</i>	Corr Coeff Sig.	,026 ,183	,058** ,002	,062** ,001	,051** ,008	,083** ,000	,054** ,004	,071** ,000		
<i>Accord Tenacity</i>	Corr Coeff Sig.	,025 ,193	,053** ,005	,073** ,000	,091** ,000	,017 ,381	,088** ,000	,026 ,175	,046* ,018	
<i>Accord Emo. Intelligence</i>	Corr Coeff Sig.	,060** ,002	,137** ,000	,102** ,000	,102** ,000	,104** ,000	,168** ,000	,029 ,127	,046* ,016	,016 ,394

Tableau B. 16 Coefficient de corrélation de Kendall pour chacune des dimensions prises deux par deux avec son seuil de significativité.

* significatif à $p < 0,05$

** significatif à $p < 0,01$

Les résultats du Tau de Kendall (Tableau B.16) vont dans le même sens que ceux obtenus lors de la comparaison du classement dimensionnel et général. Bien qu'il y ait de l'accord d'une échelle à l'autre entre les catégories dans lesquelles les observateurs sont classés, une part de différence demeure.

Cependant, il ne faut pas oublier que ces catégories d'accord sont issues d'une variable d'intervalle. Aussi, il serait intéressant de regarder au niveau des observateurs et non pas au niveau moyen par groupe.

En effet, il se peut que le désaccord soit dû à un observateur spécifique présentant une caractéristique particulière (comme son utilisation atypique de l'échelle), ou bien à un ensemble d'écarts, moindres, entre tous les observateurs.

Nous allons donc nous intéresser aux écarts, à savoir aux distances séparant chacun des membres du groupe d'observateur évaluant le même self de chacun des autres membres de son groupe. Ainsi, nous pourrions caractériser chaque observateur avec un score représentant son écart aux autres membres et nous intéresser aux caractéristiques permettant d'expliquer la variabilité de cet écart d'un observateur à un autre.

Dans ce chapitre nous nous sommes intéressés à l'accord et à la consistance inter-observateurs. Tout d'abord nous avons étudié les méthodes classiques d'évaluation et d'estimation de l'accord inter-juges avec les corrélations Bravais-Pearson et les coefficients de corrélation Intraclasse. En appliquant ces méthodes à nos résultats issu d'un questionnaire à 360°, nous avons pu constater un accord faible entre les observateurs d'un même self, et ce, qu'il s'agisse des corrélations Bravais-Pearson ou des corrélations Intraclasse. Au vu de ces résultats, nous avons adopté une autre conceptualisation de l'accord inter-observateurs – l'hypothèse de restriction de la variance. LeBreton et ses collaborateurs ont proposé dans les années 2000 d'expliquer autrement les faibles valeurs trouvées par la plupart des chercheurs pour l'accord inter-observateurs. D'après eux, ces faibles coefficients seraient dus à la non adéquation des méthodes corréationnelles pour appréhender l'accord dans le cadre d'évaluations telles que le 360°. En effet, dû fait de sélections successives mais aussi des tendances à la surévaluation dans les 360, il se produirait un phénomène de restriction de la variance qui aurait pour conséquence de diminuer les coefficients estimant l'accord entre les observateurs, si ces coefficients sont basés sur une méthode corréationnelle. Aussi, ils proposent d'utiliser d'autres indices – non corréationnels – pour estimer l'accord entre les observateurs les r_{WGS} . Ces indices permettent de comparer la variance observée au sein de chaque groupe d'observateur (variance Intra-self) avec une variance théorique, estimant la variance obtenue si les réponses des observateurs étaient aléatoires.

Dans la seconde partie de ce chapitre, nous avons donc calculé les indices r_{WG} et avons pu constater – comme d'autres études – que l'accord entre les observateurs, sur la base de ces statistiques, apparait bien meilleur. Pourtant, lors de ces analyses, nous avons également mis à jour des limites spécifiques à ces nouveaux indices comme la remise systématique à 0 des coefficients Hors-norme ou le choix de la distribution aléatoire de référence. Malgré ces difficultés, nous avons tout de même pu catégoriser l'ensemble de nos groupes d'observateurs en trois catégories d'accord sur la base des coefficients r_{WG} : les observateurs en Accord, les observateurs en Désaccord et les observateurs présentant des coefficients Hors-norme.

Nous intéressant à la stabilité de cet accord d'une dimension à une autre, nous avons été confronté au problème de la généralisabilité de l'accord. En effet, dans leurs travaux, les chercheurs calculent l'accord entre observateurs échelles par échelles (donc dans notre cas, indépendamment pour chacune des 10 dimensions du GELI). Nous avons donc voulu à apprécier la stabilité de cette catégorisation d'une dimension à l'autre mais aussi la possibilité de calculer l'accord de manière générale, sans utiliser une approche

dimension par dimension. Les résultats de cette analyse sont apparus mitigés.

Aussi, souhaitant dépasser cette situation, nous avons décidé de nous intéresser aux écarts entre les observateurs. En effet, pour mieux comprendre le désaccord, mais aussi proposer des explications de celui-ci, l'approche par catégorisation des observateurs en groupe d'accord a trouvé ses limites. Nous allons donc, dans le prochain chapitre nous étudier ce qui distingue les observateurs présentant des écarts élevés avec les autres observateurs de leurs groupes des observateurs présentant peu d'écart (étant en accord avec les autres observateurs de leur groupe).

Chapitre 7 : Étude de la variable « Écart » – synthèse des différences inter-évaluateurs.

1. Les écarts inter-observateur dans leurs évaluations

Toujours dans notre optique de mieux comprendre l'accord et le désaccord au sein des observateurs, nous avons calculé pour chaque observateur son écart moyen, avec chacun des autres observateurs de son groupe et ce, item par item et échelle par échelle.

Ainsi, le score différence en *Visioning*, correspond à l'écart moyen entre les observateurs d'un même self, moyenné sur les 8 items de l'échelle. Ce score s'exprime donc dans la même métrique que les réponses aux items.

	N	Minimum	Maximum	Moyenne	Écart-type
Écart <i>Visioning</i>	17127	,15	4,85	1,3	0,46

Tableau C.1 Statistiques descriptives de la variables Écart pour la dimension *Visioning*.

Ainsi, pour la dimensions *Visioning*, l'écart moyen varie de pas d'écart à près de cinq pont d'écart par item en moyenne pour certains observateurs. La tendance centrale est à 1,3 points d'écart, ce qui est loin d'être négligeable sur une échelle de Lickert en sept points.

Pour les autres dimensions, les résultats sont similaires comme cela peut être lu dans le tableau C.2 :

	Minimum	Maximum	Moyenne	Écart-type
<i>Visioning</i>	0,15	4,85	1,3	0,46
<i>Empowering</i>	0,2	4,05	1,26	0,42
<i>Energizing</i>	0	4,98	1,16	0,42
<i>Designing</i>	0	4,29	1,15	0,42
<i>Rewarding</i>	0,13	4,38	1,25	0,46
<i>Team Building</i>	0,09	4,29	1,26	0,41
<i>Outside Orientation</i>	0,1	5,16	1,1	0,43
<i>Global Mindset</i>	0,08	5,32	1,18	0,41
<i>Tenacity</i>	0,04	4,83	1,04	0,44
<i>Emotional Intelligence</i>	0,16	4,43	1,25	0,44

Tableau C.2 Statistiques descriptives des 10 variables « Écart » pour les 10 dimensions du GELI.

Nous pouvons remarquer que certaines échelles sont plus sensibles aux divergences que d'autres, ainsi l'écart moyen varie de un point d'échelle pour la dimension *Tenacity* à 1.26 pour les dimensions *Empowering* et *Teambuilding*.

Toutes les échelles confondues, pour un item donné, nous nous apercevons qu'en moyenne, les observateurs ont un écart dans leurs évaluations (par rapport aux autres observateurs évaluant le même self) de un point d'échelle.

	N	Minimum	Maximum	Moyenne	Écart-type
Écart Total	17127	,20	3,74	1,2	0,3

Tableau C.3 Statistiques description de la variables Écart Total

Au niveau d'ensemble, il existe donc une forte variabilité dans les écarts avec des observateurs, toutes échelles confondues, qui s'éloignent de près de quatre points en moyenne par item des autres observateurs évaluant le même self et des observateurs qui sont dans l'ensemble très proches des autres observateurs de leur groupe.

Qu'en est-il donc des groupes d'accord que nous avons constitués précédemment ? Basés sur les variances moyennes par groupe, permettent-ils aussi de saisir la variabilité entre les observateurs ? Sinon, qu'en est-il des caractéristiques classiques que nous avons étudiées pour mieux comprendre la variance des scores, ces caractéristiques

permettent-elles d'expliquer les différences de ces écarts ? En d'autres mots, certaines caractéristiques des selves et des observateurs permettent-elles de comprendre pourquoi certains observateurs sont d'accord avec les autres observateurs de leur groupe, et d'autres pas ?

2. Approche multi-niveaux des écarts entre observateurs au niveau des dimensions du GELI.

Dans la mesure où nous nous intéressons aux observateurs, il y a toujours un phénomène de dépendance au sein de nos données. Nous allons donc de nouveau recourir aux analyses multi-niveaux pour essayer d'expliquer, d'une part, la variance des écarts entre groupes d'observateurs et, d'autre part, la variance des écarts au sein des groupes d'observateurs.

Voyons d'abord si l'approche multi-niveaux se justifie pour nos données. Pour ce faire, nous pouvons vérifier la réduction de la déviance qu'on obtient pour un modèle tenant compte de la structure imbriquée comparativement au modèle avec un seul niveau.

Prenons l'exemple de la dimension *Visioning* :

$$\text{Ecart VIS}_i = \beta_0 + e_i$$

$$-2\loglikelihood = 21702,721$$

Équation C. 1 Modèle de niveau 1 et sa déviance pour la dimension Écart du facteur *Visioning*.

Pour le modèle sans multi-niveaux, nous avons donc une déviance de 21703 -2loglikelihood (Équation C.1). Par contre, pour le modèle emboîté, nous obtenons une déviance de 17625 (Équation C.2)

$$\text{Visioning}_{ij} = \beta_{0j} + e_{ij}$$

$$-2\loglikelihood = 17625,282$$

Équation C. 2 Modèle de niveau 2 et sa déviance pour la dimension Écart du facteur *Visioning*.

Donc une réduction de la déviance de $21703 - 17625 = 4078$, ce qui, avec un degré de liberté, est significatif à $P < 0.001$, montant donc l'importance d'utiliser le multi-niveaux pour appréhender ces résultats.

Nous pouvons à présent nous intéresser à la répartition de la variance entre la part de la variance totale liée à la variance inter groupe (i.e. Inter-self) et celle due aux variations inter observateurs (Intra-self).

$$Visioning_{ij} = \beta_{0j} + e_{ij}$$

$$\beta_{0j} = 1,312(0,005) + u_{0j}$$

$$u_{0j} \sim N(0, \sigma^2_{u0}) \quad \sigma^2_{u0} = 0,100(0,003)$$

$$e_{ij} \sim N(0, \sigma^2_e) \quad \sigma^2_e = 0,111(0,001)$$

$$-2 \text{ Loglikelihood} = 17625,289$$

Équation C. 3 Variance Inter- et Variance Intra-groupe pris en compte par le modèle pour la dimension Écart du facteur *Visioning*.

Toujours pour la dimension *Visioning*, la variance Inter-groupes (entre les selves), estimée par σ^2_{u0} est égale à 0.1 et la variance Intra-groupe (entre les observateurs pour chaque self), estimée par σ^2_e , est égale à 0.111. Ainsi la variance totale (σ^2_T) est donc égale à 0.211 (Équation C.4).

$$\sigma^2_T = 0,1 + 0,111 = 0,211$$

Équation C. 4 Variance totale estimée pour la dimension Écart du facteur *Visioning*.

$$VPC = \frac{\sigma^2_u}{(\sigma^2_u + \sigma^2_e)} = \frac{0,1}{0,211} = 0,47$$

Équation C. 5 *Variance Partition Coefficient*

Le VPC est donc égal à 0.47 (Équation C.5), ce qui signifie que la répartition de la variance entre l'inter et l'intra-groupe est presque équivalente. De ce fait, l'ajout au modèle de variables explicatives aussi bien au niveau II (au niveau des Self) qu'au niveau I (au niveau des observateurs) va être particulièrement intéressant.

La répartition de la variance entre Inter et Intra-groupe est similaire pour les autres dimensions du GELI (Tableau C.4).

	Pourcentage INTER-Groupe	Pourcentage INTRA-Groupe
<i>Visioning</i>	53	47
<i>Empowering</i>	47	53
<i>Energizing</i>	47	53
<i>Designing</i>	49	51
<i>Rewarding</i>	44	56
<i>Team Building</i>	46	54
<i>Outside Orientation</i>	46	54
<i>Global Mindset</i>	45	55
<i>Tenacity</i>	46	54
<i>Emotional Intelligence</i>	48	52

Tableau C. 4 Répartition de la variance en pourcentage entre l'Inter- et l'Intra-groupe pour les 10 dimensions du GELI.

Nous allons donc à présent vérifier le pouvoir explicatif des catégories de LeBreton sur la variance observée.

A. La catégorie d'accord

La catégorie d'accord a été obtenue lors des analyses précédentes en étudiant la dispersion Intra-self, Inter-observateur. Nous allons à présent vérifier que cette caractéristique permet d'appréhender une part importante de la variabilité inter-groupes, et peut-être aussi intra-groupe.

Lors de l'ajout de cette variable, la catégorie « Désaccord » sera utilisée comme catégorie de référence.

$$Ecart\ VIS_{ij} = \beta_{0j} + -0,475(0,010) Accord_j + 0,375(0,016) Hors - Norme_j + e_{ij}$$

$$-2loglikelihood = 13762,661$$

Équation C. 6 Paramètres pour les caractéristiques d'accord dans le modèle sur les écarts de la dimension *Visioning*.

Le résultat de l'équation C.6 nous montre des valeurs importantes et significatives pour les coefficients des groupes Accord et Hors-Norme. En effet, pour la dimension *Visioning*, le groupe en accord présente des écarts moyens significativement inférieurs aux deux autres

groupes. Alors que le groupe des Hors-norme, présente des écarts significativement supérieurs aux deux autres groupes (Équation C.7).

$$\text{Accord} : -\frac{0,475}{0,01} = -47,5 \quad \text{Hors-Norme} : \frac{0,365}{0,016} = 22,8$$

Équation C. 7 Valeurs des paramètres pour la catégorie Accord et Hors-norme

	Coefficient groupe Accord	Coefficient groupe Hors-Norme
<i>Visioning</i>	-0,48 *	0,38 *
<i>Empowering</i>	-0,38 *	0,33 *
<i>Energizing</i>	-0,43 *	0,31 *
<i>Designing</i>	-0,44 *	0,34 *
<i>Rewarding</i>	-0,4 *	0,35 *
<i>Team Building</i>	-0,54 *	0,35 *
<i>Outside Orientation</i>	-0,45 *	0,34 *
<i>Global Mindset</i>	-0,4 *	0,32 *
<i>Tenacity</i>	-0,49 *	0,33 *
<i>Emotional Intelligence</i>	-0,56 *	0,34 *

Tableau C.5 Coefficients pour les groupes d'accord (Désaccord et Hors-Norme) dans le modèle multi-niveaux tenant compte des groupes d'accord.

* significatif à $p < 0,001$

Pour l'ensemble des échelles du GELI, nous pouvons observer la même tendance (Tableau C. 5) : le groupe des observateurs en accord présente des écarts significativement plus faibles que les deux autres groupes et le groupe des observateurs Hors-Norme présente des écarts significativement plus importants que les autres groupes.

Ces coefficients valident donc la logique de création de ces groupes, tout en mettant en lumière l'intérêt du groupe Hors-norme. En effet, les observateurs de ce groupe présentent des écarts encore plus importants, et donc des désaccords plus importants, que les observateurs du groupe Désaccord et cela semble être leur caractéristique. Il se peut donc qu'il s'agisse de groupes d'observateurs en désaccord systématique tel qu'évoqué par LeBreton.

Grâce aux analyses multi-niveaux, nous pouvons également nous intéresser à la variance expliquée par ces catégories d'accord.

$$Ecr\ VIS_{ij} = \beta_{0j} + \beta_1 Accord_j + \beta_2 Hors-Norme_j + e_{ij}$$

$$\sigma^2_{u0} = 0,027 (0,001)$$

$$\sigma^2_e = 0,111 (0,001)$$

Équation C. 8 Variances inter- et intra-groupe du modèle tenant compte de la catégorie d'accord pour la dimension Écart du facteur *Visioning*.

Nous pouvons aussi calculer la nouvelle variance totale σ^2_T (Équation C. 9).

$$\sigma^2_T = 0,027 + 0,111 = 0,138$$

Équation C. 9 Variance Totale.

Si l'on regarde la réduction de la variance totale, nous notons qu'une part importante de variance totale a été expliquée par l'ajout de ces catégories (Équation C. 10).

$$\frac{(0,211 - 0,138)}{0,211} = 0,35$$

Équation C. 10 Part de la variance totale d'origine expliquée par les catégories d'accord.

Ainsi, 35% de la variance d'origine a été expliquée par les catégories d'accord de LeBreton. Cependant, nous pouvons remarquer que la variance expliquée est quasiment exclusivement de la variance Inter-groupe (entre les selves).

$$\frac{(0,1 - 0,027)}{0,1} = 0,73$$

Équation C. 11 Part de la variance Inter-self expliquée par le modèle.

L'équation C.11 nous montre donc que 73% de la variabilité entre les groupes de selves a pu être expliquée par les catégories de LeBreton, alors que toute la variance Intra-self est restée, elle, inexpliquée.

	Variance Totale expliquée par les catégories d'accord	Variance Inter-Groupe expliquée par les catégories d'accord
<i>Visioning</i>	34,60%	73,00%
<i>Empowering</i>	38,64%	81,93%
<i>Energizing</i>	35,63%	73,26%
<i>Designing</i>	32,78%	73,42%
<i>Rewarding</i>	36,67%	80,21%
<i>Team Building</i>	25,86%	55,00%
<i>Outside Orientation</i>	30,57%	67,82%
<i>Global Mindset</i>	31,95%	70,51%
<i>Tenacity</i>	31,44%	65,59%
<i>Emotional Intelligence</i>	27,84%	58,70%

Tableau C. 6 Pourcentage de la variance totale et de la variance inter-groupe expliqué par les catégories d'accord.

Deux conclusions peuvent être tirées de ces résultats : tout d'abord les catégories obtenues grâce aux indices r_{WG} sont associés aux écarts existant entre observateurs. Ainsi, ces écarts entre observateurs d'un même groupe sont plus importants pour les groupes en désaccord et hors-norme que pour les groupes en accord.

De même, ces catégories permettent d'expliquer entre 55 et 82% de la variabilité inter-self en termes de désaccord, ce sont donc des indicateurs utiles pour identifier les groupes d'observateur en accord et en désaccord. D'autre part, les groupes hors-norme présentent des spécificités qui les distinguent significativement des autres groupes d'observateur, et en particulier des groupes d'observateur en désaccord.

Cependant, malgré leur intérêt, ces catégories d'accord ne permettent pas d'expliquer les différences (variance) observées entre les observateurs d'un même self. Nous allons donc utiliser les caractéristiques des selves et des observateurs pour tenter de les expliquer.

2. Les variables explicatives classiques

Nous allons nous intéresser au pouvoir explicatif des 5 variables que nous avons déjà utilisé mieux comprendre les différences dans les scores des observateurs : l'âge et le genre (aussi bien du self que des observateurs) et la catégorie d'observateur (supérieur, collaborateur, subordonné ou autre).

1. Les catégories d'observateur

Nous allons dans un premier temps regarder si les différentes catégories d'observateur présentent des écarts différents – certaines catégories sont elles plus en accord que d'autres ?

La catégorie Supérieur est prise en tant que catégorie de référence

Pour la dimension *Visioning* :

$$\text{Ecart } VIS_{ij} = \beta_{0j} + \beta_1 \text{Other}_{ij} + \beta_2 \text{Co-Worker}_{ij} + \beta_3 \text{Direct Report}_{ij} + e_{ij}$$

$$-2\text{Loglikelihood} = 17598,565$$

Équation C. 12 Modèle pour la dimension Écart du facteur Visioning tenant compte des catégories d'observateurs.

Nous pouvons noter que l'ajout du paramètre catégorie d'observateur permet une réduction significative de la déviance (-2loglikelihood) à $P < 0,001$, avec 3 degrés de liberté (Équation C.13).

$$17625,3 - 17598,6 = 26,7$$

Équation C. 13 Réduction de la déviance entre le modèle nul et le modèle de l'équation C.12

Pour les autres dimensions du GELI une réduction de la déviance peut aussi être observée (Tableau C. 7).

	Différence
<i>Visioning</i>	26,7 ***
<i>Empowering</i>	4,1
<i>Energizing</i>	69,5 ***
<i>Designing</i>	13,1 **
<i>Rewarding</i>	132,9 ***
<i>Team Building</i>	40,2 ***
<i>Outside Orientation</i>	60,5 ***
<i>Global Mindset</i>	30,9 ***
<i>Tenacity</i>	24,2 ***
<i>Emotional Intelligence</i>	26,6 ***

Tableau C. 7 Réduction de la déviance (-2 Loglikelihood) entre le modèle nul et le modèle tenant compte des catégories d'observateur.

** significatif à $p < 0,01$

*** significatif à $p < 0,001$

Seule la dimension *Empowering* ne présente pas une réduction de la déviance significative suite à l'ajout de la variable catégorie d'observateur. Pour cette dimension, il semble que la catégorie d'observateur ne permette pas de mieux comprendre la variabilité des écarts. De même, la réduction la plus importante est observée pour la dimension *Rewarding and Feedback*, pour laquelle la catégorie d'observateur joue un rôle important dans les écarts observés entre observateurs.

Nous pouvons à présent nous étudier les coefficients des différentes catégories d'observateurs.

$$\text{Ecart VIS}_{ij} = \beta_{0j} + -0,001(0,011) \text{Other}_{ij} + 0,028(0,01) \text{Co-Worker}_{ij} + 0,036(0,009) \text{Direct Report}_{ij} + e_{ij}$$

Équation C. 14 Paramètres associés au modèle de la dimension Écart du facteur *Visioning*.

Pour les collaborateurs et les subordonnés, les coefficients sont significatifs (Équation C. 15). Ainsi, les écarts entre les observateurs, pour la dimension *Visioning*, tendent à être plus importants pour les Collaborateurs et les Subordonnés que pour les Supérieurs ou les observateurs de la catégorie Autres.

$$\text{Collaborateur: } \frac{0,028}{0,01} = 2,8 \quad \text{Subordonné: } \frac{0,036}{0,009} = 4$$

Équations C. 15 Valeurs des paramètres pour la catégorie Collègue et Subordonné

Pour les autres dimensions du GELI les paramètres sont quelque peu différents, bien qu'on retrouve les mêmes tendances (Tableau C. 8).

	Collaborateur	Subordonné	Autre
<i>Visioning</i>	0,03 **	0,04 ***	0
<i>Empowering</i>	0,01	0,02	0
<i>Energizing</i>	0,05 ***	0,04 ***	-0,01
<i>Designing</i>	0,02	0,02	-0,01
<i>Rewarding</i>	0,03 ***	0,09 ***	0,01
<i>Team Building</i>	0,04 ***	0,03 ***	0
<i>Outside Orientation</i>	0,03 **	0	-0,04 ***
<i>Global Mindset</i>	0	-0,02 **	-0,04 ***
<i>Tenacity</i>	0,02 *	-0,01	-0,02
<i>Emotional Intelligence</i>	0,04 ***	0,03 ***	0,01

Tableau C.8 Coefficients pour les catégories d'observateurs dans le modèle multi-niveaux tenant compte des catégories d'observateurs.

* significatif à $p < 0,05$

** significatif à $p < 0,01$

*** significatif à $p < 0,001$

Les dimensions *Visioning*, *Energizing*, *Rewarding*, *Team Building* et *Emotional Intelligence* présentent les mêmes tendances, à savoir que les écarts entre les observateurs tendent à être plus importants pour les observateurs qui sont Collaborateurs et Subordonnés que pour les observateurs qui sont Supérieurs ou autres.

Pour les dimensions *Outside Orientation*, *Global Mindset* et *Tenacity*, les observateurs catégorisés comme autres et dans une moindre mesure les Subordonnés, présentent des écarts plus faibles dans leur évaluations comparativement aux Supérieurs. Alors que les collaborateurs présentent, eux, les écarts les plus importants surtout pour les dimensions *Outside orientation* et *Tenacity*.

Pour *Empowering*, la réduction de la déviance ayant été non significative, il n'est pas étonnant que les coefficients ne soient pas significatifs.

En conclusion, nous pouvons dire que les observateurs de la catégorie Autre sont ceux qui présentent, toutes échelles confondues, les écarts les plus faibles, ces écarts sont similaires pour les observateurs de la catégorie Supérieur. Pour les collaborateurs et les subordonnés mis à part pour 2 dimensions (*Global Mindset* et *Tenacity*), ils présentent des écarts plus importants que les observateurs de la catégorie Supérieur.

2. L'âge

Nous allons à présent regarder dans quelle mesure l'âge des observateurs et des selves permet de comprendre les écarts observés.

$$\text{Ecart } VIS_{ij} = \beta_0 + \beta_1 \hat{\text{Age}}_{OBS_{ij}} + \beta_2 \hat{\text{Age}}_{SELF_j} + e_{ij}$$

$$-2 \text{Loglikelihood} = 17453,337$$

Équation C.15bis Modèle pour la dimension Écart du facteur *Visioning* tenant compte des âges des observateurs et des selves.

Pour la dimension *Visioning*, la réduction de la déviance observée après l'ajout des variables âge des observateurs et âge des selves est significative à $P < 0,001$ avec 2 degrés de liberté (Équation C. 16).

$$17625,3 - 17453,3 = 172$$

Équation C.16 Réduction de la déviance entre le modèle nul et le modèle tenant compte des âges des observateurs et du self pour la dimension la dimension Écart du *Visioning*.

Les résultats sont similaires pour les autres dimensions du GELI (Tableau C. 9).

	Différence
<i>Visioning</i>	171,95 **
<i>Empowering</i>	112,45 **
<i>Energizing</i>	65,55 **
<i>Designing</i>	73,31 **
<i>Rewarding</i>	128,06 **
<i>Team Building</i>	110,56 **
<i>Outside Orientation</i>	111,7 **
<i>Global Mindset</i>	101,38 **
<i>Tenacity</i>	109,76 **
<i>Emotional Intelligence</i>	91,96 **

Tableau C. 9 Réduction de la déviance (-2 Loglikelihood) entre le modèle nul et le modèle tenant compte des âges des observateurs et des selves.

** *significatif à $p < 0,001$*

L'ajout de ces deux variables explicatives « âge » permet une réduction significative de la déviance pour toutes les échelles du GELI.

Cependant, aucun des deux paramètres (âge du self ; âge de l'observateur) n'est significatif sur les dix échelles du GELI. Nous pouvons donc conclure que ni l'âge des observateurs, ni l'âge du self que les observateurs évaluent, n'a un impact direct sur les écarts (le désaccord) entre les évaluateurs.

Nous allons à présent nous intéresser à la dernière variable – le genre.

3. Le genre

Tout comme pour l'âge, nous allons nous intéresser aussi bien au genre du self évalué qu'à celui des observateurs qui l'évaluent. Nous allons donc chercher à comprendre si les écarts existant entre les observateurs d'un même self peuvent être expliqués par du genre.

Pour l'ensemble des variables nous utiliserons la modalité homme comme référence.

$$\text{Ecart } VIS_{ij} = \beta_{0j} + \beta_1 \text{GenreOBS:Female}_{ij} + \beta_2 \text{GenreSELF:Female}_j + e_{ij}$$

$$-2\text{Loglikelihood} = 17376,006$$

Équation C. 17 Modèle pour la dimension Écart du facteur *Visioning* tenant compte du genre des observateurs et du genre du self.

Tout d'abord nous pouvons noter que l'ajout de ces deux variables « genre » permet une réduction de la déviance pour la dimension *Visioning*, significative à $P < 0,001$ avec deux degrés de liberté (Équation C.18).

$$17625,3 - 17376 = 249,3$$

Équation C. 18 Réduction de la déviance entre le modèle nul et le modèle tenant compte des genres des observateurs et du self pour la dimension Écart du facteur *Visioning*.

Pour l'ensemble des autres dimensions du GELI, l'ajout de ces variables permet également une réduction significative de la déviance (Tableau C.10).

	Différence
<i>Visioning</i>	249,3 **
<i>Empowering</i>	178,2 **
<i>Energizing</i>	178,2 **
<i>Designing</i>	237,8 **
<i>Rewarding</i>	257,9 **
<i>Team Building</i>	254,2 **
<i>Outside Orientation</i>	169,7 **
<i>Global Mindset</i>	150,6 **
<i>Tenacity</i>	140,1 **
<i>Emotional Intelligence</i>	240,1 **

Tableau C. 10 Réduction de la déviance (-2 Loglikelihood) entre le modèle nul et le modèle tenant compte du genre des observateurs et des selves.

** significatif à $p < 0,001$

Comme le modèle avec les variables explicatives Genre du self et Genre de l'observateur permet une réduction significative de la déviance pour toutes les échelles du GELI, nous pouvons à présent nous intéresser aux paramètres des deux catégories « genre ».

Pour la dimension *Visioning* :

$$\begin{aligned} \text{Ecart } VIS_{ij} = & \beta_{0j} + 0,030(0,007) \text{GenreOBS:Female}_{ij} \\ & + 0,075(0,013) \text{GenreSELF:Female}_j + e_{ij} \end{aligned}$$

Équation C. 19 Paramètres du modèle tenant compte des genres des observateurs et du self pour la dimension Écart du facteur *Visioning*.

En observant les coefficients, nous remarquons que les observatrices femmes tendent à avoir des écarts significativement plus importants que les observateurs hommes, et que les selves femme, tendent à avoir des observateurs plus en désaccord que les selves hommes (Équation C.20).

$$\text{Genre observateurs: } \frac{0,03}{0,007} = 4,3 \quad \text{Genre Selves: } \frac{0,075}{0,013} = 5,8$$

Équation C. 20 Valeur des paramètres pour les variables Genre des observateurs et Genre des selves.

Pourtant, comme nous pouvons le voir dans le tableau C. 11, ces résultats ne sont pas identiques pour toutes les dimensions du GELI.

	Observateur femme	Self femme
<i>Visioning</i>	0,03 ***	0,08 ***
<i>Empowering</i>	0,02 ***	0,02 *
<i>Energizing</i>	0,03 ***	0
<i>Designing</i>	0,04 ***	-0,01
<i>Rewarding</i>	0,04 ***	0
<i>Team Building</i>	0,04 ***	0,01
<i>Outside Orientation</i>	0	-0,02
<i>Global Mindset</i>	0	-0,01
<i>Tenacity</i>	0,02 **	0
<i>Emotional Intelligence</i>	0,03 ***	0,02

Tableau C.11 Coefficients pour le genre dans le modèle multi-niveaux tenant compte du genre des selves et des observateurs.

* significatif à $p < 0,05$

** significatif à $p < 0,01$

*** significatif à $p < 0,001$

Pour les observateurs tout d'abord, nous pouvons remarquer que huit des dix dimensions (*Outside Orientation* et *Global Mindset*), présentent des résultats similaires. En effet, pour ces dimensions, les observatrices femmes tendent à avoir plus d'écart que les observateurs hommes, donc à être significativement plus en désaccord lors de l'évaluation du même manager. Ce résultat peut être mis en lien avec nos résultats obtenus dans la partie précédente, où nous avons conclu que les femmes donnaient des scores significativement plus élevés que les hommes, aux selves qu'elles évaluaient.

Le genre du self évalué, a un effet sur les écarts entre les observateurs, pour deux dimensions seulement : *Visioning* et *Empowering*. Pour ces deux dimensions, les observateurs évaluant des selves femmes tendent à être plus en désaccord et à avoir donc des écarts plus importants que ceux évaluant des selves hommes. Cela se comprend au regard de nos résultats précédents. En effet, les selves femmes tendent à recevoir des scores plus élevés que les selves hommes – à l'exception de ces deux dimensions *Visioning* et *Empowering*. Sur ces deux dimensions, et en particulier sur la dimension *Visioning*, les femmes reçoivent des scores plus faibles que les selves hommes. Cela peut donc en partie expliquer la présence d'écarts significatifs.

Aussi, bien que toutes ces variables explicatives permettent des réductions de la déviance significatives, il serait intéressant de voir la part de variance que l'ensemble de ces variables permettent d'expliquer.

4. Le modèle complet

En prenant toujours l'exemple de la dimension *Visioning* nous allons nous intéresser au modèle complet, pour savoir quelle part de la variance ces cinq caractéristiques permettent d'expliquer.

$$\begin{aligned} \text{Ecart } VIS_{ij} = & \beta_{0j} + \beta_1 \text{Other}_{ij} + \beta_2 \text{Co-Worker}_{ij} + \beta_3 \text{Direct Report}_{ij} + \beta_4 \text{ÂgeOBS}_{ij} \\ & + \beta_5 \text{ÂgeSELF}_j + \beta_6 \text{GenderOBS:Female}_{ij} + \beta_7 \text{GenderSELF:Female}_j + e_{ij} \end{aligned}$$

$$-2 \text{Loglikelihood} = 17258,716$$

Équation C. 21 Modèle général pour la dimension Écart du facteur *Visioning* tenant compte de la catégorie d'observateur, des âges et des genres des observateurs et du genre du self.

L'ajout des cinq variables explicatives dans le modèle permet une réduction de la déviance significative à $P < 0,001$, avec 7 degrés de liberté (Équation C. 22).

$$17625 - 17258 = 367$$

Équation C. 22 Réduction de la déviance entre le modèle nul et le modèle tenant compte de la catégorie d'observateur, des âges et des genres des observateurs et du genre du self.

Nous pouvons aussi, nous intéresser à la variance expliquée par l'ensemble de ces cinq variables. Pour ce faire nous devons tout d'abord calculer $\sigma^2_{T_model}$ (Équation C. 23) puis la comparer avec la variance totale du modèle d'origine σ^2_T .

$$\sigma^2_{T_model} = 0,098 + 0,11 = 0,208$$

$$\frac{(\sigma^2_T - \sigma^2_{T_model})}{\sigma^2_T} = \frac{(0,211 - 0,208)}{0,211} = 0,014$$

Équation C. 23 Par de variance totale d'origine expliquée par le modèle.

L'ajout de ces cinq variables nous ont donc permis d'expliquer 1,4% de la variance totale d'origine des écarts observés entre les observateurs sur la dimension *Visioning*.

Nous pouvons voir dans le tableau C. 12, que ces cinq variables explicatives permettent une réduction significative de la déviance pour l'ensemble des échelles du GELI, bien que la part de variance qu'elles expliquent est faible.

	Réduction de la déviance	Part de variance expliquée
Visioning	367 *	1,40%
Empowering	235 *	0,60%
Energizing	280 *	0,00%
Designing	286 *	0,60%
Rewarding	415 *	1,00%
Team Building	350 *	1,20%
Outside Orientation	311 *	1,00%
Global Mindset	226 *	0,10%
Tenacity	226 *	0,00%
Emotional Intelligence	302 *	0,50%

Tableau C. 12 Réduction de la déviance (-2loglikelihood) entre le modèle nul et le modèle avec les 5 variables explicatives et part de variance libre expliquée par les 5 variables explicatives ajoutées.

* *significatif à $p < 0,001$ avec 7 degrés de liberté*

Ces résultats sont quelque peu décevant vu l'ampleur de la variance dont nous disposons et que nous souhaitons expliquer. Pourtant, à la différence des scores aux échelles que nous ne pouvions étudier qu'individuellement, l'écart entre les observateurs peut également être conçu globalement.

Cette approche nous permettra peut-être de voir des tendances générales, non visibles au niveau de chacune des dimensions.

C. L'étude des écarts au niveau général.

Le niveau général d'écart à été construit en moyennant sur les échelles les écarts obtenus pour chacun des observateurs.

Comme nous l'avons fait lors de notre étude des dimensions du GELI, nous allons regarder quelle part de la variance ces cinq caractéristiques permettent d'expliquer.

En d'autres mots, nous allons chercher au niveau général, dans quelle mesure ces variables explicatives nous permettraient de réduire la variabilité dans les écarts des évaluateurs, donc dans leur désaccord.

Pour ce faire nous allons dans un premier temps construire un modèle nul (Équation C. 24).

$$\begin{aligned}
 \text{Ecart } TOTAL_{ij} &= \beta_{0j} + e_{ij} \\
 \beta_{0j} &= 1,201(0,003) + u_{0j} \\
 u_{0j} &\sim N(0, \sigma^2_{u0}) \quad \sigma^2_{u0} = 0,049(0,001) \\
 e_{ij} &\sim N(0, \sigma^2_e) \quad \sigma^2_e = 0,043(0,001) \\
 -2 \text{ Loglikelihood} &= 2383,19
 \end{aligned}$$

Équation C. 24 Variance Inter- et Variance Intra-groupe du modèle pour la dimension Écart Total.

1. Le modèle général

La variable Écart Total a été calculée en effectuant une moyenne de tous les écarts entre les observateurs d'un même self pris deux par deux et ce sur l'ensemble de dimensions du GELI.

$$\begin{aligned}
 \text{Ecart } TOTAL_{ij} &= \beta_{0j} + -0,019(0,007) \text{ Other}_{ij} + 0,021(0,006) \text{ Co-Worker}_{ij} \\
 &+ 0,013(0,006) \text{ Direct Report}_{ij} + -0,001(0,000) \hat{\text{Age}}_{OBS_{ij}} + -0,001(0,000) \hat{\text{Age}}_{SELF_j} \\
 &+ 0,022(0,004) \text{ Gender}_{OBS:Female_{ij}} + 0,010(0,009) \text{ Gender}_{SELF:Female_j} + e_{ij} \\
 -2 \text{ Loglikelihood} &= 2233,887
 \end{aligned}$$

Équation C. 25 Modèle général pour la dimension Écart Total tenant compte de la catégorie d'observateur, des âges et des genres des observateurs et du genre du self.

Nous pouvons noter tout d'abord une réduction significative de la déviance qui était de 2382.2 pour le modèle nul et qui passe à 2233.9 pour le modèle avec les cinq variables explicatives (Équation C.25). Cette réduction de 148 est significative avec 7 degrés de liberté.

En ce qui concerne les paramètres des caractéristiques explicatives, certaines conclusions énoncées au niveau des dimensions sont confirmées.

Ainsi, l'âge, que ce soit au niveau des selves ou des observateurs n'est pas lié aux écarts existants entre les évaluations des observateurs d'un même self.

Pour la catégorie d'observateur, au niveau général, nous tirons des conclusions similaires à celles de la dimension *Visioning*. Tout d'abord les observateurs Autres présentent les écarts – désaccords – les plus faibles. Ensuite, les Collaborateurs et les Subordonnés présentent, eux, les désaccords les plus importants, ce qui les différencie significativement aussi bien des observateurs de la catégorie Autre, que des Supérieurs.

$$\begin{aligned} \text{Collaborateur: } \frac{0,021}{0,006} &= 3,5 & \text{Subordonné: } \frac{0,013}{0,006} &= 2,2 \\ \text{Autres: } -\frac{0,019}{0,007} &= -2,7 \end{aligned}$$

Équation C. 26 Valeur des paramètres pour les variables Collaborateur, Subordonné et Autre.

Pour le genre, nous confirmons également la conclusion tirée sur les dimensions individuelles. Les femmes observatrices tendent à avoir des écarts – désaccords – significativement plus importants que les observateurs hommes. Alors que le genre du self évalué, au niveau général, n'affecte pas significativement l'ampleur des écarts.

$$\text{Genre observateurs: } \frac{0,022}{0,004} = 5,5 \quad \text{Genre Selves: } \frac{0,010}{0,009} = 1,1$$

Équation C. 27 Valeur des paramètres pour les variables Genre des observateurs et Genre des selves.

Enfin, nous allons à présent nous intéresser à l'ampleur de la variance expliquée par toutes ces variables.

Nous nous apercevons dès lors que la variance restée libre dans le modèle σ_{T_Model} est identique à la variance totale du modèle nul σ_T (Équation C. 28). L'ensemble de ces cinq variables ne nous a donc pas permis d'expliquer une part visible de la variance.

$$\begin{aligned} \sigma_{T_model} &= 0,049 + 0,043 = 0,092 \\ \sigma_T &= 0,049 + 0,043 = 0,092 \end{aligned}$$

Équation C. 28 Part de la variance d'origine expliquée par les cinq variables explicatives.

Lors de notre étude du désaccord et plus spécifiquement des groupes d'accord construit selon la méthode de LeBreton, nous avons noté que ces groupes présentaient une utilisation différente de l'échelle. Aussi, nous avons souhaité vérifier si l'utilisation de

l'échelles, et plus spécifiquement de l'ancrage « deux » pouvait nous permettre d'expliquer une part de variance plus importante.

2. La fréquence d'utilisation de l'ancrage 2

Comme nous l'avons montré dans les analyses descriptives faites en première partie (Chapitre 6), les observateurs en désaccord tendent à être ceux, utilisant l'ensemble de l'échelle et surtout les valeurs faibles. Aussi, nous allons ajouter une nouvelles variable explicative « le nombre de 2 utilisé dans le questionnaire en moyenne ».

$$Ecart\ TOTAL_{ij} = \beta_{0j} + 2,542(0,027) Nombre\ 2 + e_{ij}$$

$$\beta_{0j} = 1,131(0,003) + u_{0j}$$

$$\sigma^2_{u0} = 0,033(0,001)$$

$$\sigma^2_e = 0,028(0,000)$$

Équation C. 29 Variance Inter- et Variance Intra-groupe du modèle pour la dimension Écart Total.

Nous voyons tout d'abord un fort coefficient positif significatif à $P < 0,001$ pour l'utilisation moyenne de l'ancrage 2. Ainsi, plus les observateurs tendent à utiliser des valeurs 2 dans leurs réponses, plus l'écart qu'ils ont avec les autres observateurs de leur groupe est élevé.

$$Nombre\ moyen\ de\ 2 : \frac{2,542}{0,027} = 94,1$$

Équation C. 30 Valeur du paramètre pour la variable Nombre moyen d'utilisation de l'ancrage 2.

Il est bien évident que cette constatation peut sembler quelque peu simpliste, mais nous tenons à attirer l'œil du lecteur sur l'importance que l'utilisation ou non de cet ancrage permet d'expliquer en terme de variabilité.

$$\text{Variabilité Inter – Self} : \frac{(0,049 - 0,033)}{0,049} = 0,33$$

$$\text{Variabilité Intra – Self} : \frac{(0,043 - 0,028)}{0,043} = 0,35$$

Équation C. 31 Part de la variance inter et intra-self expliqué par la variable « Nombre d'utilisations moyennes de l'ancrage 2 »

Nous voyons aussi que l'ajout de cette variable nous permet d'expliquer 33% de la variance Inter-self et 35% de la variance Intra-self, Inter-observateur (Équation C.31).

Comparativement aux faibles parts de variances expliquées précédemment, nous pouvons voir ici qu'une part importante de la variance existante est expliquée par l'utilisation de l'échelle.

Nous nous sommes intéressés dans ce chapitre à la variable « Écart ». Cette variable correspond de l'écart moyen, par groupe, entre tous les observateurs pris deux par deux. Elle résume donc les différences inter-observateur, intra-self. Comme cela pouvait être attendu, nous avons grâce aux coefficients d'accord de LeBreton, pu expliquer une grande part de la variance de cette variable – surtout en ce qui concerne la variabilité inter-groupe donc entre les selves. Cela n'est guère surprenant sachant que les coefficients r_{WG} sont construits sur la base de la dispersion des observateurs au sein des groupes de selves. Ce qui est plus intéressant en revanche, c'est la spécificité du groupe Hors-Norme. Nous avons montré que ce groupe présente des caractéristiques différentes des groupes en désaccord. En effet, il présente des coefficients significativement supérieurs aussi bien aux groupes en accord qu'aux groupes en désaccord. Les groupes d'observateurs présentant des coefficients r_{WG} Hors-Norme présentent donc significativement plus d'écarts entre eux dans leurs évaluations du même self que les observateurs des autres groupes.

Pourtant, la majeure partie de la variabilité entre les observateurs au sein des groupes de selves reste inexpliqué. Nous avons donc étudié le pouvoir explicatif des variables classiques (catégorie d'observateur, âge et genre) sur l'écart entre les observateurs. Cependant, avec l'ensemble de ces cinq caractéristiques, seule une part très faible de la variance a pu être expliquée. Nous avons alors introduit une nouvelle variable explicative « l'utilisation de l'ancrage 2 », pour étudier l'impact et l'effet de l'utilisation différente de l'échelle sur la variance observée. Cette dernière variable explicative a permis, à elle seule, d'expliquer une part importante de la variance. Il semble donc que les écarts observés entre les évaluateurs d'un même self, sont principalement reliés à l'utilisation de l'échelle et ne peuvent être expliqués par des variables socio-démographiques classiques.

Du fait de ces résultats, une question plus théorique sur le 360 se trouve posée. Maintenant que nous avons démontré la part très importante de variance libre présente dans les évaluations à 360°, quelle base psychométrique peut-on appliquer à ces données ? Quel modèle de la mesure permettrait de tenir compte de cette variabilité inhérente ?

Chapitre 8 : Le score vrai et la théorie de la mesure

Nous en arrivons donc à une conclusion quelque peu limitée, à savoir que la majeure partie des différences observées entre les observateurs d'un même self ne peuvent être expliquées par aucune des variables classiquement utilisées dans les études sur les 360.

Ce constat nous fait revenir à la question fondamentale et méthodologique de la mesure au sens psychométrique du terme. Pour introduire la démarche mise en œuvre dans ce chapitre il importe de rappeler deux éléments théoriques.

D'une part, que la mesure en psychologie a pris naissance avec les tests d'intelligence et a été formalisée sous le nom de théorie psychométrique de la mesure. Cette théorie psychométrique a ensuite été élargie à l'évaluation des aptitudes en général, puis a été appliquée aux questionnaires dans le domaine de l'évaluation de la personnalité.

D'autre part, que les approches à 360° sont principalement issues du courant managérial et de gestion. Elles visaient au départ à étendre les évaluations classiques réalisées en entreprises (le plus souvent sous forme de *feedback* unidirectionnel des supérieurs vers les subordonnés) vers des évaluations multiples. Aussi, le 360 avait pour objectif affiché de permettre une prise de conscience des managers et surtout des *top-managers* de la perception que leur supérieur hiérarchique, leurs collègues et leurs subordonnés avaient de leurs comportements et de leur management.

Les psychométriciens ne se sont à proprement parler intéressés aux évaluations à 360° que dans un second temps, lorsqu'ils ont commencé à adapter des questionnaires en « version 360 », permettant donc de mettre au regard de l'auto-évaluation, une hétéro-évaluation par des observateurs, ayant chacun un rapport particulier et différent au self évalué.

Du point de vue des 360°, leur méthode vise deux objectifs : d'une part permettre une évaluation plus exacte (en ne se limitant plus à la simple auto-évaluation) et, d'autre part, faire prendre conscience au manager de points de vue différents (ses supérieurs d'une part, mais aussi ses collègues, ses subordonnés, ses clients et autres).

Du point de vue psychométrique, la question cruciale qui se pose est de savoir comment, dans cette approche plurielle de la mesure, on peut identifier une « mesure vraie ». On pourrait à la limite se demander si l'hétéro-évaluation et l'intérêt porté aux différences entre évaluateurs n'est pas en contradiction avec la théorie du score vrai qui est par essence, unique.

1. Le score vrai et la théorie de la mesure.

A. La théorie classique

La psychométrie repose sur la théorie du score vrai et de l'erreur de mesure. Le score vrai d'un individu suppose que lorsque l'on s'intéresse à une aptitude donnée (aptitude mathématique, aptitude spatiale...) chaque individu peut être situé précisément sur l'échelle latente de la dimension d'intérêt en fonction du degré plus ou moins élevé auquel il possède cette aptitude.

Un des premiers psychométriciens à mettre clairement en forme cette approche fût Gulliksen en 1950, dans son livre *Theory of Mental Tests*. Ce livre proposa pour la première fois une théorie des scores aux tests, formulée mathématiquement.

Cette théorie définit le score vrai d'une personne comme la somme de son score observé plus ou moins l'erreur de mesure.

$$\text{Score Observé} = \text{Score vrai} + \text{Erreur de Mesure}$$

Équation D. 1 : Score vrai.

Nous introduirons la théorie des tests à partir de la présentation de Laveault et Grégoire (1998) qui résumant la théorie classique des scores en sept postulats :

- Le premier postulat reprend notre formule de l'équation D.1, en précisant que l'erreur de mesure se distribuant normalement, le score observé se distribuera donc lui aussi normalement autour du score vrai.

- Le second postulat dérive du précédent et note que le score vrai d'un individu est l'espérance mathématique des scores observés, s'il y a plusieurs passations indépendantes du test au même individu. De fait, si l'on pouvait administrer de manière indépendante le même test au même individu un nombre « infini » de fois, la moyenne des scores observés serait le score vrai. Ainsi, la dispersion des scores observés autour du score vrai constitue l'erreur type de mesure pour cet individu.
- Le troisième postulat précise qu'il n'y a pas de corrélation entre l'erreur de mesure et le score vrai. De fait, il n'y a pas de mécanisme permettant de réduire ou d'augmenter la part d'erreur en fonction de la compétence réelle du sujet.
- Le quatrième postulat dit que les erreurs de mesure à deux tests différents ne sont pas corrélées entre elles.
- Le cinquième postulat précise d'ailleurs qu'il n'y a pas non plus de corrélation entre le score vrai à un test et l'erreur de mesure à un autre test.
- Le sixième postulat, pose que deux tests ne pourront être considérés comme parallèles que si et seulement si leurs scores vrai et leurs erreurs de mesure sont identiques.
- Enfin, le septième postulat précise que deux tests sont dits tau-équivalents quand leurs scores vrais ne diffèrent que par une constante additive.

Cette théorie classique des tests considère donc l'erreur comme strictement aléatoire, de sorte que les diverses sources d'erreur sont indépendantes et vont s'annuler en présence d'un nombre répété de mesures.

Mais si l'erreur n'est plus aléatoire – les postulats précédents ne sont plus valables. Dans ce cas, cette erreur est considérée comme systématique. Ainsi la formule finale peut-être présentée comme l'équation D. 2.

Score observé = score vrai + erreur systématique + erreur aléatoire.

Équation D. 2 : Décomposition du score observé

Nous voyons donc après l'énumération de ces postulats le lien qu'entretient la théorie de la mesure avec la fidélité d'un test et sa validité et donc l'approche psychométrique au sens large.

En effet, la logique même de la fidélité est dérivée de la théorie classique, les variances étant également réparties en variance vraie et variance d'erreur, nous pouvons définir la fidélité d'un test en tant que la proportion de sa part de variance vraie par rapport à sa part de variance d'erreur.

Ainsi la SEM (*Standard Error of Measurement*) est l'erreur présente dans le score de la personne testée, et elle est inversement proportionnelle à la fidélité (Equation D. 3).

$$s_E = s_x \sqrt{1 - r_{xx}}$$

Équation D. 3: L'erreur standard de mesure

où s_x est l'écart-type de la variable x et r_{xx} la fidélité de la variable x .

D'autre part, la validité peut être à son tour définie comme la part de score vrai dénuée d'erreur systématique.

Il est intéressant de noter, tel que cité par Gulliksen dans son livre, que Spearman en 1904 et 1907 proposait déjà toutes les formules qui seront utilisées plus tard dans la théorie du score vrai.

Ainsi Spearman, au début du vingtième siècle, travaillant sur l'intelligence et les aptitudes, a montré le lien entre la longueur du test, la longueur du critère utilisé pour sa validation et la validité de l'outil, donc le score vrai. Proposant par exemple la Prophétie de Spearman-Brown, permettant de corriger a-posteriori les coefficients de fidélité selon la longueur du test. Sa préoccupation était donc bien de démontrer la part de score vrai dans le score final qui est, lui, la somme des items du test.

B. L'estimation ponctuelle du score vrai

A partir de ce que nous venons de rappeler sur la théorie des scores, nous voyons que la question qui se pose pour le 360 est : « quel est le score vrai de la personne ? »

Selon les concepteurs des 360, l'un des objectifs premiers de ce type d'évaluation est de mettre en exergue les perceptions (moyennes) différentes de chacune des catégories d'observateur selon leur situation hiérarchique par rapport au manager évalué.

Les différences observées entre ces groupes d'évaluateurs deviennent donc, lors de l'interprétation des résultats, des différences « vraies » porteuses de sens. Or si tel est le cas, nous serions donc portés à conclure qu'il y a autant de scores vrais qu'il y a d'observateurs – ou de groupes d'observateurs.

Pourtant une telle conclusion pose de sérieux problèmes psychométriques, non seulement de validation mais aussi d'interprétation. Qui plus est, nous avons vu dans les analyses précédentes (Chapitre 7) que seule une part très faible des différences entre les observateurs peut être imputée à leur catégorie d'appartenance et que la majeure part de ces différences reste inexplicée même après prise en compte de toutes les caractéristiques socio-économiques classiques.

Dans le cadre du 360, l'estimation ponctuelle du score vrai pourrait donc résulter de l'agrégation des scores des observateurs. Mais si l'on agrège les scores des observateurs (même par catégories hiérarchiques: supérieur, subordonné, etc.), il faut justifier de la cohérence entre les observateurs sinon ce score agrégé ne sera pas consistant. Or, les valeurs des ICC que nous avons présenté au chapitre 5, ne supportent pas cette agrégation car elles présentent des valeurs inférieures à .70, or .70 est souvent jugée comme la valeur au-delà de laquelle on peut parler d'un consensus inter-juges fiable.

De même, lors de nos analyses des réponses des observateurs, nous avons pu mettre en évidence l'utilisation particulière que certains d'entre eux avaient de l'échelle de réponse (échelle de Lickert en sept points).

Au niveau théorique, comme nous l'avons vu avec Spearman, il est important de garder à l'esprit que la théorie classique des tests a été proposée pour les mesures d'aptitude, et qu'elle a été ensuite transposée aux mesures de personnalité. Ces dernières présentent pourtant une différence majeure : il ne s'agit plus dans les questionnaires de personnalité, d'évaluer la performance à travers des réussites ou des échecs, mais d'évaluer la position d'un individu sur un trait latent grâce à ses réponses à un questionnaire d'auto-évaluation.

Malgré cette différence de fond, la théorie n'a pas été modifiée. Par contre, de plus en plus de recherches ont été mises en place pour tenter de pallier au mieux les limites de l'auto-évaluation (auto- et hétéro-duperie, tendance à répondre vers la moyenne, effets parasites de facteurs situationnels : Friedman & Amoo, 1999; Watkins & Cheung, 1995 ; pour une revue récente - Vrignaud & Loarer, 2008).

1. Les biais dans l'utilisation des échelles de Lickert – les ancrages

Parmi les biais classiquement associés aux échelles de Lickert, l'un des premiers concerne le choix des adverbes et adjectifs d'ancrage.

En effet, outre le problème lié à la compréhension des termes indiquant la fréquence avec laquelle un comportement est effectué tels que « Toujours » ou « Parfois », plusieurs études ont montré que selon leur groupe d'appartenance (femmes au foyer, étudiants, professionnels) les adjectifs et adverbes utilisés comme ancrage n'étaient pas nécessairement compris de la même manière (Myers & Warner, 1968).

Aussi, en demandant à des sujets d'associer un nombre de jours par mois avec plusieurs adverbes de fréquence (assez souvent, plutôt souvent et pas très souvent) Bradburn et Miles (1979) ont pu montrer une très grande variabilité dans les réponses. De même, Bartram et Yielding (1973) ont trouvé une plus grande consistance entre les répondants pour les termes positifs (tels que « Extrêmement »¹⁷ ou « Totalement »¹⁸) que négatifs (comme « Pas vraiment »¹⁹).

Concernant les termes positifs et négatifs justement, une tendance des répondants à vouloir assigner des valeurs plutôt positives que négatives a également été montrée.

17 Extremely

18 Totally

19 Not quite

Enfin, la transformation en nombres des adjectifs et adverbes (par exemple 1 – Jamais ; 2 – Parfois, etc.) peut également changer le sens de ces mots pour certains répondants. Ainsi, Schwarz et ses collaborateurs (1991) ont posé la question suivante à des répondants adultes: « À quel point avez-vous réussi dans la vie ? ²⁰ ». Ils ont trouvé que 34% des répondants choisissaient des réponses se rangeant entre 0 et 5 lorsque l'échelle de réponse allait de 0 « Pas réussi du tout » à 10 « Réussite exceptionnelle ». Alors que seuls 13% des répondants choisissaient des réponses entre -5 et 0 si l'échelle allait de -5 « Pas réussi du tout » à 5 « Réussite exceptionnelle ».

Connaissant les difficultés pouvant être liées aux ancrages dans les questionnaires de type Lickert et au vu des résultats d'analyse que nous avons obtenu sur l'utilisation de l'échelle par les observateurs, nous pourrions envisager une intervention au niveau de l'échelle de réponse. Celle-ci viserait à améliorer l'accord inter-juges et, de la sorte, nous permettre de construire une estimation ponctuelle du score vrai du manager. La technique des vignettes d'ancrages pourrait, de ce fait, s'avérer une méthode intéressante pour le 360.

2. La technique des Vignettes d'ancrage

La technique des Vignettes d'ancrage (*Anchoring Vignettes* - King, Murray, Salomon, & Tandon, 2004; Wand, 2007) se propose de décrire des individus ou des situations représentant l'un, puis l'autre pôle pour chacune des dimensions mesurées. Ces vignettes hypothétiques doivent être évaluées par le répondant sur la même échelle de Lickert que dans le reste du questionnaire.

Ainsi, dans le cadre du GELI, pour la dimension *Visioning*, nous pourrions imaginer une première vignette décrivant un manager présentant tous les comportements de cette dimension, et ce fréquemment :

- ✧ Le Manager A, manager dans la société N. s'efforce de saisir toutes les opportunités se présentant à lui et n'a pas peur de remettre en cause le *statut quo* existant dans l'entreprise. Il cherche systématiquement à avoir une vue de

20 « How successful would you say you have been in life? »

l'ensemble de la situation et parvient à simplifier celles qui seraient trop complexes.
Ce comportement fait de Monsieur A. un bon leader.

Une seconde vignette porterait sur un manager présentant ces comportements moins fréquemment :

- ✧ Le Manager B, manager dans la société N. s'efforce de saisir les opportunités se présentant à lui et s'autorise à remettre en cause le *statut quo* existant dans l'entreprise lorsque nécessaire. Il cherche à avoir une vue d'ensemble des situations et à simplifier les situations complexes.
Ce comportement fait de Monsieur B. un bon leader.

Et enfin une dernière vignette présentant ces comportements rarement :

- ✧ Le Manager C, manager dans la société N. cherche à saisir les opportunités se présentant à lui sans pour autant remettre en cause le *statut quo* existant dans l'entreprise. Il parvient à avoir une vue d'ensemble des situations, bien qu'il lui soit difficile de simplifier les situations complexes.
Ce comportement fait de Monsieur C. un bon leader.

Le répondant doit évaluer sur une échelle de Lickert, dans quelle mesure il est en accord ou en désaccord avec le fait que le comportement de chacun de ces managers de la société N fait d'eux un bon leaders.

Pour ce qui est de la méthode de comparaison, le score corrigé C_i est dérivé du score y_i tel qu'indiqué par l'évaluateur sur un item du questionnaire. L'intérêt est que la mesure non paramétrique de la variable sous-jacente, peut être comparée entre des répondants différents.

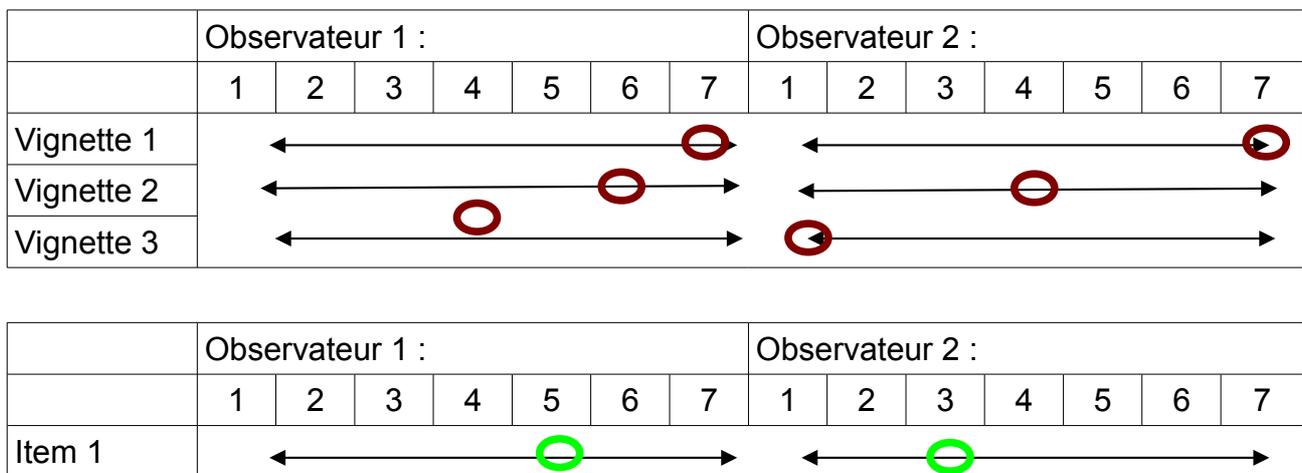
En effet, la réponse d'un répondant donnée est re-codée selon que celle-ci est plus élevée, égale ou inférieure au niveau auquel il a évalué les vignettes d'ancrage (J). De la sorte, cette méthode transforme l'échelle de réponse en catégories de réponse où le nombre de catégories finales correspond à $2J + 1$ et ces catégories sont comparables d'un individu à l'autre quel que soit leurs références pour l'ancrage.

Ainsi, pour notre exemple nous aurions $2 \times 3 + 1$ catégories de réponses. Ces nouvelles catégories de réponse, pour notre exemple seraient donc :

- La réponse à l'item 1 < réponse à la Vignette 3
- La réponse à l'item 1 = réponse à la Vignette 3
- La réponse à l'item 1 est comprise entre la réponse à la Vignette 3 et la Vignette 2
- La réponse à l'item 1 = réponse à la Vignette 2
- La réponse à l'item 1 est comprise entre la réponse à la Vignette 2 et la Vignette 1
- La réponse à l'item 1 = réponse à la Vignette 1
- La réponse à l'item 1 > réponse à la Vignette 1

Prenons deux observateurs :

Pour les vignettes d'ancrage, l'observateur 1 a décrit le Manager A à 7 sur 7, le manager B à 6 sur 7 et le manager C à 4 sur 7. L'observateur 2 a décrit le Manager A à 7 sur 7, le manager B à 4 sur 7 et le manager C à 1 sur 7 (Graphique Z. 1).



Graphique Z.1 : Score aux vignettes d'ancrage et à un item du questionnaire de deux observateurs.

Pour le questionnaire à proprement parlé, l'observateur 1 a attribué une note de 5, alors que l'observateur 2 une note de 3 à l'item un (Graphique Z. 1).

Après re-codage, nous pouvons remarquer dans le graphique Z. 2 que la note de l'observateur 2 reste inchangé du fait de ses réponses aux vignettes d'ancrage et reste donc à 3, alors que celle de l'Observateur 2 est diminuée de 5 à 3.

En comparant ces observateurs, nous nous apercevons alors que ceux-ci sont en fait en accord.

						
1	2	3	4	5	6	7
$i1 < V3$	$i1 = V3$	$V3 < i1 < V2$	$i1 = V2$	$V2 < i1 < V1$	$i1 = V1$	$i1 > V1$

Graphique Z. 2: Score re-codé des deux observateurs.

Cette technique des vignette d'ancrage est une méthode qui s'est largement répandue ces dix dernières années, surtout dans le domaine de la santé et des sondages sur les effets des diverses politiques économiques (King et al., 2004), par contre elle reste très rare dans le domaine de la psychologie (Möttus et al., 2012).

On peut cependant s'interroger sur l'apport de cette méthode à des données telles que celles du GELI dans la mesure où, comme nous l'avons constaté, une majorité des répondants utilise seulement les échelons les plus à droite et quelques-uns répartissent leurs réponses sur une plus grande étendue des échelons.

Ainsi, si les répondants les plus extrêmes à droite, cotent les vignettes en utilisant toute l'étendue de l'échelle (ce qui serait cohérent avec la description des « vignettes »), en utilisant les échelons 1, 2 ou même 3 pour la vignette basse, les échelons 4 et 5 pour la vignette moyenne et les échelons 6 et 7 pour la vignette haute – la plupart de leurs réponses seront re-codées dans la catégorie haute.

Les répondants ayant utilisé toute l'étendue de l'échelle pour évaluer le leader et ayant répondu de manière similaire aux vignettes, auront leurs réponses re-codées qui couvrent toute l'étendue de l'échelle.

L'écart entre ces deux types de répondants se retrouvera donc largement dans le re-codage, présentant les mêmes difficultés que celles que nous avons observées dans notre travail. Néanmoins, il serait intéressant de mettre en œuvre cette méthode qui permettrait d'évaluer l'impact de l'ancrage sur ces différences observées entre répondants.

Ainsi, si comme notre conclusion le suggère, la méthode des vignettes ne nous permettrait pas forcément de rapprocher les évaluations des observateurs, une autre approche peut être envisagée pour une estimation ponctuelle du score vrai.

Nous pourrions en effet, prendre comme valeur ponctuelle, l'estimation de l'observateur qui serait le plus fiable.

C. Le modèle du consensus culturel

Si la construction d'un « score vrai » en agrégeant les réponses des observateurs ne semble pas offrir toutes les garanties de fiabilité, on pourrait s'interroger sur le fait que certains observateurs sont plus fiables que d'autres.

Le modèle du consensus culturel est une approche qui s'inscrit dans les recherches ethnologiques et anthropologiques. Dans le cadre de ces recherches, les anthropologues sont souvent confrontés aux difficultés de recueil de données, mais surtout aux difficultés liées aux désaccords entre les informateurs. Ces désaccords entre les informateurs donnent naissance à deux problèmes d'une part, comment estimer les « connaissances culturelles » des informateurs et, d'autre part, comment est-il possible d'inférer, avec un degré de confiance suffisant, la réponse « correcte » à une question donnée.

C'est dans le cadre de cette problématique que s'inscrit le modèle de consensus culturel développé par Romney, Batchelder et Batchelder (1986). Cette approche vise à estimer la compétence des informateurs tout en supposant que les bonnes réponses ne sont pas connues a priori. Les bonnes réponses s'entendent ici au sens des réponses le plus largement partagées par les informateurs, c'est à dire reflétant à un consensus culturel dans le groupe des informateurs.

Ainsi nous voyons clairement le parallèle pouvant être fait avec les 360. Dans les évaluations à 360, nous avons également un ensemble d'informateurs qui donnent des réponses différentes sur une question dont nous ne connaissons pas la réponse vraie (le score vrai du manager qu'ils évaluent sur la dimension concernée).

Ainsi, nous pourrions appliquer l'approche de Romney et Batchelder, qui proposent de considérer comme le plus compétent (le plus informé), l'observateur dont les réponses sont les plus proches de celles de l'ensemble des autres observateurs, et d'associer un poids aux évaluations des observateurs en fonction de leur « compétences ».

Cette approche, visant à obtenir une estimation ponctuelle unique semble cependant présenter des limites si elle est appliquée aux résultats des 360. Tout d'abord, cette méthode s'applique dans le cas où il y a effectivement une bonne réponse. Ce qui est vrai lorsque la question d'intérêt est « Comment appelez-vous cet objet », mais est plus discutable pour le domaine de la personnalité, et plus encore du 360 où l'expérience d'un subordonné avec son manager n'est pas nécessairement plus ou moins exacte que l'expérience d'un collaborateur avec le même manager. Dans le cadre d'une évaluation de type 360 il est donc difficile de parler d'un consensus culturel.

Ces remarques tendent à mettre plus encore en lumière les limites d'une approche du score vrai en termes d'estimation ponctuelle. En effet, pour l'approche psychométrique classique, toutes les différences observées autour du score vrai correspondraient à de l'erreur. Or, nous pourrions également utiliser ces différentes évaluations des observateurs d'un même self pour construire la distribution des scores du self. Ainsi, le score vrai ne serait pas uniquement l'estimation ponctuelle d'une valeur mais la construction d'une distribution

De plus, dans les 360 les différences entre les observateurs ne sont pas considérées systématiquement comme de l'erreur aléatoire : les différences entre les catégories d'observateurs sont associées à des différences de contexte, qu'il est nécessaire d'intégrer pour avoir une évaluation « vraie » de la personne. Ainsi, en s'appuyant uniquement sur la théorie psychométrique classique du score vrai, toutes les différences observées dans chaque groupe self-observateurs ne seraient imputables qu'à de l'erreur de mesure. L'interprétation de ces écarts deviendrait donc dénuée de pertinence.

Aussi, nous allons considérer à présent une approche différente du score vrai - en abordant les évaluations faites par les observateurs en termes de distribution.

2. L'estimation du score vrai sous forme de distribution

A. La valeur manquante fondamentale de Rubin

Les évolutions les plus importantes de la théorie du score vrai se sont développées dans le cadre des modèles de réponse à l'item et plus particulièrement en éduométrie (c'est à dire dans le cadre des enquêtes sur l'évaluation des compétences des élèves). Le statisticien David Rubin (1987) a proposé une théorie des valeurs manquantes dans laquelle il classe ces valeurs en trois catégories selon les causes à l'origine de leur absence dans les données (*Missing Completely at Random*, *Missing at Random* et *Missing Not at Random*). Il propose ensuite un ensemble de traitements statistiques pouvant être réalisés avec des données présentant ces valeurs manquantes, selon la catégorie à laquelle cette absence appartient.

Dans le cadre de son travail, il a introduit une nouvelle conception du score sur la dimension latente (score vrai de la personne). En effet, pour lui, la valeur manquante fondamentale n'est pas tant l'absence de réponse à un item mais la position du sujet sur le trait latent (donc le score vrai du sujet au sens de la psychométrie). D'où l'idée d'appliquer les méthodes utilisées pour le traitement des valeurs manquantes pour estimer le score vrai. L'estimation de celui-ci peut-être conçue comme l'imputation d'une valeur manquante conditionnellement à l'information que l'on possède sur le sujet (ses réponses aux items et ses autres caractéristiques).

La méthode d'imputation multiple de Rubin (1987) pour traiter ces informations manquantes a l'avantage de pallier les limites de l'imputation simple qui proposait de remplacer l'élément manquant en utilisant soit la moyenne conditionnelle ou inconditionnelle soit une valeur de la distribution conditionnelle ou inconditionnelle. Tout en gardant les avantages de cette méthode comme la création, après imputation, d'une série de donnée complète, pouvant être analysée avec les méthodes statistiques classiques.

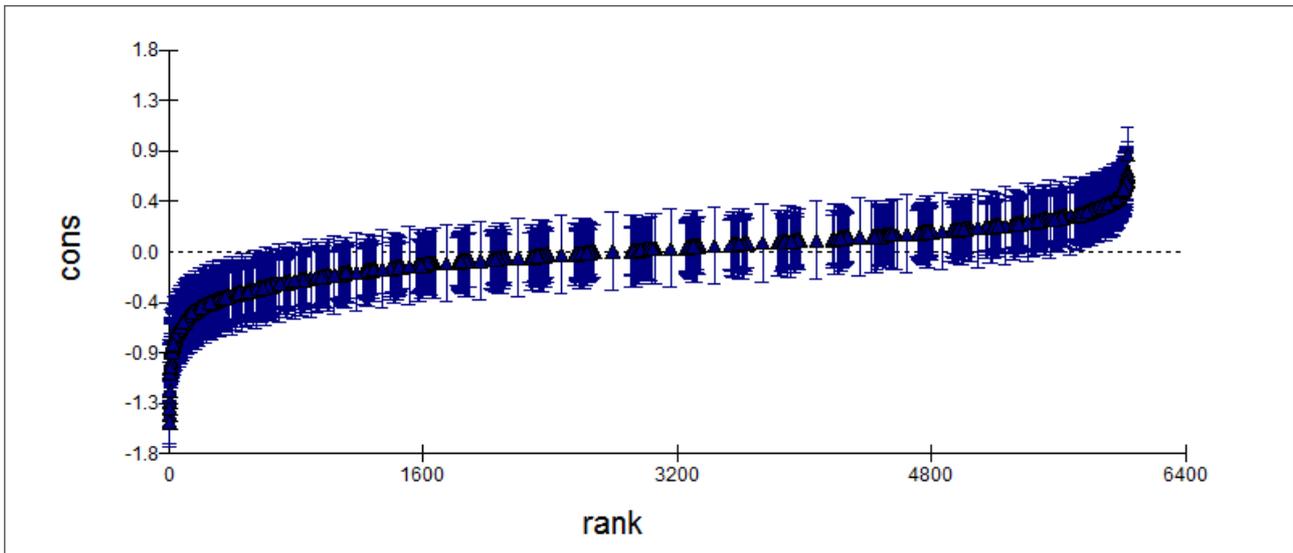
L'élément central de l'approche d'imputation multiple de Rubin est la construction d'une distribution de valeurs plausibles pour la valeur manquante. Chacune d'entre elles permettant de définir autant de séries complètes de données différentes. Les résultats (qui

peuvent varier) sont ensuite combinés par des techniques statistiques simples pour obtenir une estimation des statistiques d'ensemble et d'erreur standard de mesure (Schafer & Graham, 2002).

Rubin a appliqué cette approche conceptuelle à l'estimation du score vrai dans le cadre des enquêtes américaines et internationales sur l'évaluation des compétences des élèves. Des algorithmes ont été implantés dans les logiciels destinés au traitement des données de ces enquêtes afin de calculer des imputations multiples (en général cinq) pour chacun des élèves participant à ces enquêtes. Cette approche modifie profondément la conception du score vrai qui n'est plus l'estimation d'une valeur ponctuelle mais la construction d'une distribution. Si cette approche est satisfaisante d'un point de vue théorique, elle n'a pas été employée en dehors du champ de l'éducatrice et même plus particulièrement des enquêtes internationales sur l'évaluation des compétences des élèves. Cela tient sans doute au fait que cette méthode est bien adaptée à l'estimation de valeurs au niveau de l'échantillon (les moyennes nationales des pays participant à ces enquêtes) et non au niveau de l'individu. L'estimation du score vrai d'un individu par plusieurs valeurs n'est pas simple d'utilisation dans la mesure où elle nécessite de calculer plusieurs statistiques (par exemple autant de corrélations que de jeux de données imputées) puis d'agréger ces valeurs (cf. supra). Cette méthode a d'ailleurs donné lieu à peu de publications en dehors de ce champ des enquêtes internationales sur les compétences des élèves (Vrignaud, 2006).

Appliquée aux évaluations à 360°, cette méthode pourrait pourtant permettre une approche intéressante des scores. En effet, en reprenant la définition de Rubin sur la variable manquante fondamentale, nous pouvons concevoir le score vrai du manager comme une variable manquante et réaliser pour celle-ci une distribution des scores plausibles sur la base des informations disponibles (évaluations des observateurs et leur variabilité). De cette distribution pourraient être tirées des valeurs plausibles (de cinq à vingt selon les auteurs ; Schafer & Graham, 2002) qui serviraient à estimer les indicateurs statistiques classiques. Par exemple si l'on veut calculer une corrélation entre la performance au test et le niveau socio-économique, la réussite de l'entreprise ou la satisfaction de l'équipe, on devra en fait calculer cinq à vingt corrélations : une avec chacune des valeurs plausibles. On fera ensuite le calcul de la moyenne et de la dispersion de ces corrélations pour réaliser des tests d'inférence.

Pour chaque manager évalué, nous pourrions donc obtenir un résultat similaire au Graphique D. 1, présentant leur score moyen associé à sa dispersion.



Graphique D.1 Résidus des évaluations faites par les observateurs pour la dimension *Visioning*

Le graphique D.1 représente les résidus, avec un intervalle de confiance de 95%, pour la dimension *Visioning*. Nous voyons ainsi que les intervalles de confiance ne varient que peu en largeur. La taille de l'intervalle de confiance d'un résidu donné dépend de l'erreur standard associé à ce groupe, qui est inversement proportionnelle à la taille du groupe. Aussi, bien que certains sels aient peu alors que d'autres ont beaucoup d'observateurs, la plupart des groupes présentent un nombre d'observateurs compris entre quatre et six. Ces résidus représentent donc l'écart de chaque groupe à la moyenne générale de la dimension *Visioning*. Ainsi, un groupe dont l'intervalle de confiance ne comprend pas la moyenne générale (la ligne zéro), diffère significativement de la moyenne au niveau de confiance de 5%.

Nous voyons sur la gauche du graphique D.1 un ensemble de groupes présentant des résidus inférieurs à la moyenne, et sur la droite un ensemble de groupes avec des résidus supérieurs à la moyenne.

Pour la dimension *Visioning* se dessine donc une distribution du score des sels, à partir de l'estimation de leur variance Intra.

Ainsi, au lieu de rechercher une estimation ponctuelle du score du manager, le 360 pourrait aboutir à une estimation de la distribution, conditionnellement à la variance intra. On s'inscrirait ainsi dans une approche bayésienne en construisant la distribution du score du sujet conditionnellement aux réponses de ses observateurs aux items. On tiendrait ainsi compte de la distribution plus ou moins dispersée de la compétence du sujet tel qu'évaluée par ses observateurs.

Bien que la distribution soit construite à partir des scores observés et que cela ne s'inscrirait plus, à proprement parler, dans la théorie du score vrai, chaque leader peut se voir affecté une distribution dont l'étendue serait fonction de la variabilité de ses observateurs (comme le montre le caterpillar plot, Graphique D. 1). Cette distribution peut être construite par l'analyse factorielle multi-niveaux à partir des réponses des observateurs aux items.

B. Analyse Factorielle Confirmatoire multi-niveaux

L'analyse factorielle se réfère à une méthode statistique visant soit à explorer soit à confirmer la structure sous-jacente d'un ensemble de variables (ou d'items) ayant des liens prédéfinis avec un construit latent. L'approche exploratoire vise à identifier l'existence de structure sous-jacente, alors que l'approche confirmatoire vise, elle, à valider une structure définie a priori par les hypothèses (Nunally & Bernstein, 1994). On l'appelle aussi modèle de mesure (*measurement model*) car elle permet de construire la variable latente mesurant le score vrai.

En ce qui concerne le leadership, beaucoup d'outils ont employé une démarche factorielle pour définir et valider les dimensions mesurées, pourtant, ce qui est plus étonnant, c'est que malgré sa structure profondément hiérarchiques et imbriquées (évaluation par des groupes d'observateurs d'un même leader) peu de recherches ont eu recours à l'analyse factorielle confirmatoire dans le cadre des modèles multi-niveaux. Il faut dire que ces approches sont relativement récentes, bien que plusieurs logiciels (LISREL version 8, Mlwin, Mplus) offrent la possibilité d'estimer les paramètres de modèles d'analyse factorielle confirmatoire pour des données multi-niveaux en séparant une estimation du modèle pour la variance intra-groupe et une part du modèle pour la variance inter-groupe. De manière schématique, cela revient à analyser d'une part la matrice de

variance/covariance intra- et d'autre part la matrice de variance/covariance inter-.

Aussi, tout comme pour les analyses de la variance, ne pas tenir compte de cette structure lors des analyses factorielles peut facilement être source de més-interprétation de liens ou de relations. Dans ses démonstrations, Muthén (1994) a montré qu'ignorer la structure multi-niveaux des données entraînait des distorsions importantes au niveau des résultats. Ainsi, que ce soit en conduisant des analyses factorielles sur l'ensemble des observations (en ignorant leur structure imbriquée), ou en moyennant les réponses aux items au niveau de chaque groupe, le résultat et la structure factorielle obtenue sera un mélange de celle existant au niveau des individus et celles existant au niveau des groupes.

1. Le modèle

Si l'on prend l'exemple de la dimension *Visioning* nous pouvons grâce à l'analyse factorielle confirmatoire multi-niveaux estimer les coefficients de corrélation Intraclasse à partir des scores sur la variable latente. Ceux-ci sont considérés comme des scores vrais car ils sont mesurés sans erreur selon la conception des modèles d'équations structurales.

En effet, comme le démontre Muthén (1991), parce que la variance a pu être décomposée aux deux niveaux d'analyse : l'intra-groupe (*W – Within*) et l'inter-groupe (*B - Between*), cette variance a pu être associée soit à de la variance vraie soit à de la variance d'erreur.

Nous obtenons donc quatre paramètres représentant l'ensemble de la variance observée (Équation D. 4).

$$\text{Var Tot} = \text{BF} + \text{BE} + \text{WF} + \text{WE}$$

Équation D.4 : Décomposition de la variance Totale

où BF est la variance vraie inter-groupe, BE la variance d'erreur inter-groupe, WF, la variance vraie intra-groupe et WE la variance d'erreur intra-groupe.

Ainsi, ce modèle implique que la corrélation entre deux individus *i* et *i'* dans le groupe *g* pour la variable y_j peut être résumé par l'équation D. 5.

$$\text{Corr} (y_{gij}, y_{gi'j}) = (\text{BF} + \text{BE}) / (\text{BF} + \text{BE} + \text{WF} + \text{WE})$$

Équation D.5 : Corrélation

La Corrélation Intraclasse correspond elle aussi, à une proportion de variance Inter-groupe sur la variance vraie totale (Équation D. 6). Aussi, plus cette corrélation sera élevée, plus les observations pourront être considérées comme indépendantes. En effet, à l'extrême, si BF et BE sont nuls pour toutes les variables, alors l'analyse multi-niveaux n'est plus utile.

La part de variance Inter-groupe dénuée d'erreur est donc une estimation non biaisée du coefficient de corrélation intraclasse pour chaque variable.

$$ICC = BF / (BF + WF)$$

Équation D.6 : Corrélation Intraclasse

Pour la dimension *Visioning* nous avons réalisé les analyses sur le logiciel MPLUS version 6 (Muthén & Muthén, 2011) et nos résultats sont présentés dans le Tableau D. 1.

Item	Valeur de ICC
1	0,21
2	0,21
3	0,19
4	0,12
5	0,14
6	0,11
7	0,13
8	0,15

Tableau D.1 Coefficients de corrélation Intraclasse pour les items de la dimension *Visioning*.

Les quelques différences qui peuvent être observées entre ces valeurs et les valeurs notées lors de la première analyse (Chapitre 5), sont d'une part liées au fait que cette analyse tient compte de la structure multi-niveaux des données, mais aussi que tous les observateurs – même si leur nombre varie d'un self à l'autre – ont tous été pris en compte.

En ce qui concerne les résultats de l'analyse factorielle confirmatoire multi-niveaux à proprement parler, nous nous attendons – au vu des qualités des items – que ces 8 items présentent certaines différences parmi les observateurs.

En effet, les items de *Visioning* visant à exprimer la tendance du leader à inspirer une vision, nous pouvons tout à fait – au vu des diverses théories évoquées dans la première partie, penser que selon le rapport que l'observateur entretient avec le leader (qu'il soit son subordonné ou son supérieur) sa perception de ses capacités à inspirer une vision ne seront pas les mêmes. D'autre part des variables comme l'âge ou le genre peuvent aussi donner des perceptions différentes de la capacité d'une personne donnée à inspirer une vision.

D'autre part, nous nous attendons également à ce que ces 8 items reflètent des différences réelles entre les managers évalués – certains présentant ces qualités de manière plus marquée que les autres. En effet, malgré l'hypothèse de restriction de la variance de LeBreton (2003) évoquée dans le chapitre 6, nous pensons tout de même que des différences subsistent entre les managers, même de haut niveau.

Ainsi les réponses aux items ne refléteront pas uniquement des différences entre des managers mais aussi des différences entre les perceptions qu'ont d'eux les observateurs. Ainsi, dans les résultats, nous pourrons comparer les saturations des items (variance vrai) et les erreurs de mesure sur le facteur latent (dans cet exemple, *Visioning*) d'une part pour les observateurs et d'autre part pour les groupes, donc pour les selves en tant que tel.

Nous pourrons ainsi décomposer la variance vrai BF, la variance d'erreur BE, la variance vraie WF et erreur WE

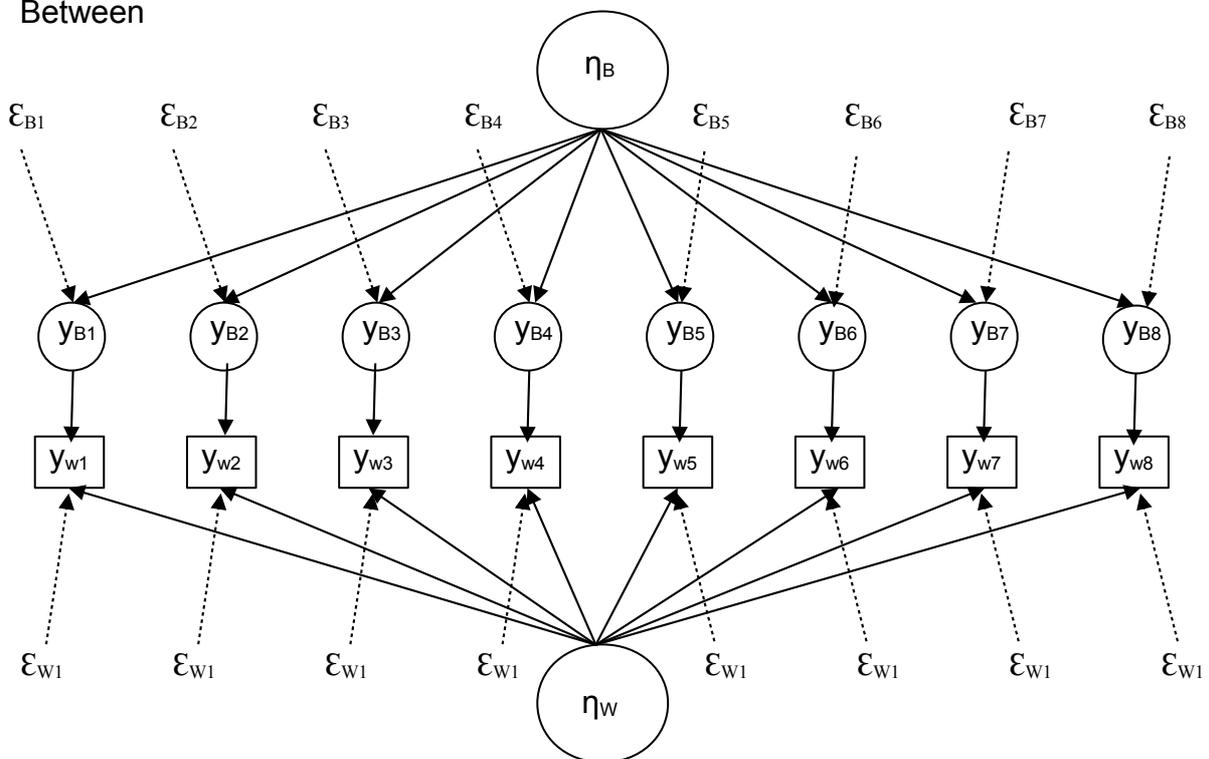
Nous avons donc construit un modèle factoriel confirmatoire tel que proposé par Muthèn (1994). Notre modèle est composé d'un facteur latent et de huit indicateurs observés (pour la dimension latente *Visioning*) représentés par les rectangles y_{w1} à y_{w8} dans le graphique D. 2. Ces indicateurs observés sont les réponses des observateurs aux 8 items de la dimension *Visioning*.

La partie inférieure de la figure (Graphique D. 2) dénommée « Intra-groupe » correspond à une analyse factorielle confirmatoire qui serait réalisée sur l'ensemble des observateurs dégroupés (pris sans tenir compte de leur appartenance locale). Chacune des 8 variables observées saturent un facteur latent (η_w) au niveau Intra-groupe et comportent, chacune, une part d'erreur aléatoire (ε_{w1} à ε_{w8}).

La partie supérieure de la figure (Graphique D. 2) est dénommée « Inter-groupe » et comporte 8 indicateurs représentés par des cercles y_{B1} à y_{B8} . Il ne s'agit pas à proprement parlé de variables observées, ces cercles représentent plutôt des moyennes pour chacune des variables observées (y_{w1} à y_{w8}). Ces « moyennes » saturent une variable latente agrégée de *Visioning* (η_B), et comportent chacune une part d'erreur (ϵ_{B1} à ϵ_{B8}).

Ainsi, comme le modèle relie les indicateurs latents agrégé et désagrégé, les valeurs observées sont donc considérées comme une fonction des construits latents inter- et intra-groupe (η_w et η_B respectivement). L'analyse factorielle confirmatoire consiste donc à analyser simultanément les matrices de covariances inter et intra.

Inter-groupe
Between



Intra-groupe
Within

Graphique D. 2 Modèle théorique pour l'analyse factorielle confirmatoire multi-niveaux

L'estimation de ces modèles donne d'une part des indicateurs d'adéquation du modèle et, d'autre part, l'estimation des saturation des facteurs, des variances des facteurs et des résidus.

2. L'adéquation du modèle

Comme pour toute analyse factorielle confirmatoire, plusieurs indices sont proposés pour estimer la qualité de l'ajustement du modèle aux données. Le « $\chi^2 - Goodness of fit$ » est l'indice le plus ancien et le plus connu. Ainsi, un modèle bien ajusté se doit d'avoir un niveau de χ^2 faible relativement au nombre de degrés de liberté. Cependant, l'effectif entre dans le calcul du χ^2 et plus la taille de l'échantillon est importante, plus le χ^2 sera élevé. De ce fait, comme les échantillons sont en général très importants dans les études ayant recours au multi-niveaux, cet indice est de moins en moins utilisé. On recommande donc de s'appuyer sur d'autres indices moins sensibles à la taille de l'échantillon comme le CFI (*Comparative Fit Index*), le RMSEA (*Root Mean Square Error of Approximation*) et le SRMR (*Standardized Root Mean Square Residual*).

Les indices obtenus pour *Visioning* sont présentés dans le Tableau D. 2.

CFI	0,92
RMSEA	0,06
SRMR (Intra-groupe)	0,04
(Inter-Groupe)	0,1

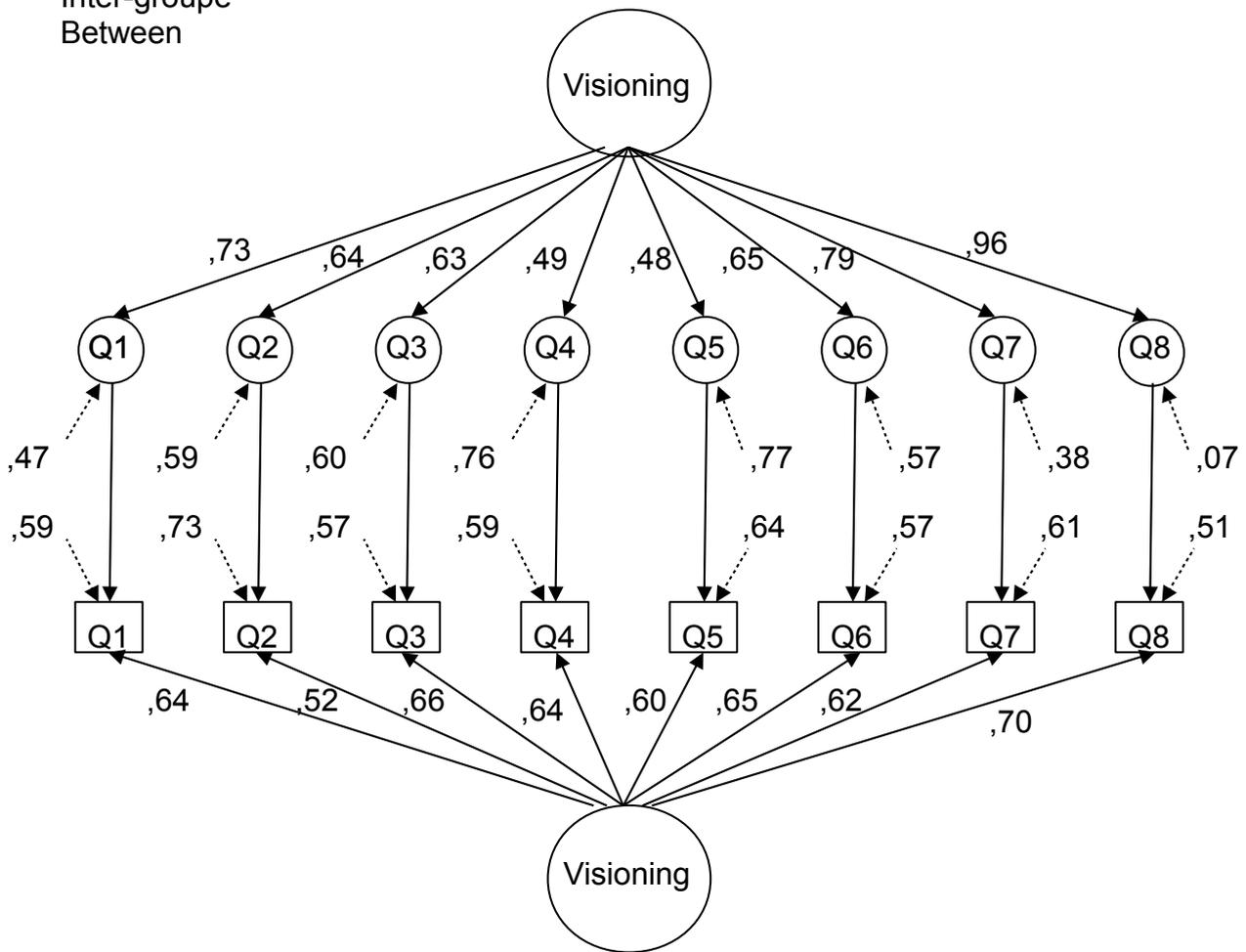
Tableau D.2 Indices CFI, RMSEA et SRMR pour la dimension *Visioning*

Ainsi, en suivant les valeurs recommandés par Hu et Bentler (1999) nous pouvons conclure à un ajustement suffisamment bon du modèle pour la dimension *Visioning* pour l'intra-groupe. Par contre l'ajustement du modèle n'est pas bon pour l'inter-groupe.

Pour les autres dimensions, nous retrouvons des résultats similaires pour les dimensions *Energizing*, *Team Building*, *Outside Orientation* et *Tenacity*. Pour les autres dimensions, l'ajustement du modèle n'est pas suffisamment bon.

Concernant les saturations, le graphique D. 3 montre les résultats que nous obtenons pour la dimension *Visioning*.

Inter-groupe
Between



Intra-groupe
Within

Graphique D. 3 Saturations du modèle factoriel confirmatoire multi-niveaux pour la dimension Visioning.

Nous voyons donc apparaître des différences d'un item à l'autre. En effet, certains items, sont plus associés à la composante latente que d'autres. Nous voyons également des différences entre le modèle Inter et Intra-groupe.

Pour le modèle Intra-groupe, les saturations sont plutôt fortes et se situent pour la plupart entre .60 et .66, l'échelle semble donc avoir une structure unidimensionnelle au niveau Intra. Les items sont donc clairement associés à la dimension *Visioning*, et reflètent la perception que les observateurs ont du manager – ce niveau d'analyse est, par ailleurs, le niveau d'intérêt principal du 360.

La structure de l'échelle est moins claire en ce qui concerne le niveau d'analyse par groupe (Inter). En effet, le modèle semble mal s'adapter aux données et les saturations des items sont très variés allant de .40 à .96 pour l'item 8 « Je donne envie à mes collaborateurs à voir au-delà des limites existantes. » Il semble donc que le modèle unidimensionnel ne soit pas le modèle le plus adapté. Aussi, si le niveau d'analyse inter-groupe était notre niveau d'intérêt principal, nous pourrions envisager de restructurer le facteur *Visioning*, en supprimant les questions 4 et 5 par exemple.

Nous pouvons aussi nous intéresser aux R², les part des variances communes entre les items et les construits latents :

	R ² Inter-groupe	R ² Intra-groupe
Item 1	0.535	0.410
Item 2	0.414	0.274
Item 3	0.400	0.434
Item 4	0.242	0.412
Item 5	0.235	0.359
Item 6	0.426	0.426
Item 7	0.625	0.388
Item 8	0.930	0.486

Tableau D. 3 Parts de variance communes entre les items et les construits latents pour les 8 items de *Visioning*.

Nous retrouvons donc la conclusion précédente concernant les items de la dimension *Visioning* : les part de variance de chacun des items expliqués par le construit latent varient d'un item à l'autre mais également selon que l'on s'intéresse au modèle Inter- ou Intra-groupe.

Nous pouvons donc conclure pour la dimension *Visioning*, qu'il semble exister pour cette dimension une structure de construit différente selon le niveau d'analyse auquel on s'intéresse. Les techniques d'analyse factorielle confirmatoire multi-niveaux proposent donc une méthodologie permettant d'explorer systématiquement la question de la structure des dimensions d'un tests.

Cependant, aujourd'hui, un nombre très limités d'études ont recours à cette méthode pour appréhender les résultats issus des questionnaires sur le leadership. Les quelques études référencées se sont, la plupart du temps, intéressés aux questionnaires d'auto-évaluation du leadership et visaient des comparaisons inter-culturelles ou inter-entreprises (Dyer, Hanges, & Hall, 2005; Hall, Hanges, & Dyer, In Press).

Il serait donc intéressant d'appliquer cette méthodologie au-delà des comparaisons culturelles, en visant comme niveau d'intérêt non plus le niveau sociétal, mais le niveau Intra-groupe qui dans le cas du 360 correspond au niveau Inter-observateurs, Intra-self.

Dans ce chapitre nous avons cherché à positionner les réponses issues des questionnaires à 360° selon l'approche psychométrique. Après avoir rappelé le modèle psychométrique classique du score vrai et de l'erreur de mesure, nous avons cherché à le mettre en relation avec le modèle des questionnaires à 360. Aussi, nous avons proposé plusieurs approches visant à permettre une estimation ponctuelle du score vrai, selon la théorie classique de la mesure (amélioration de l'accord inter-observateurs, approche par consensus culturel). Pourtant, nous avons rapidement fait face à une inadéquation centrale entre le modèle psychométrique qui considère que toute différence observée entre les scores est assimilable à de l'erreur de mesure, et les tenants des évaluations à 360° qui cherchent à interpréter les écarts existant entre les groupes d'observateurs.

Partant de ce constat, nous avons choisi de nous tourner vers une approche de la mesure quelque peu différente, qui propose de considérer le score vrai non plus de manière ponctuelle, mais sous forme de distribution.

Dans le cadre de cette approche nous sommes partis du modèle de Rubin (1987) qui considère que la valeur manquante fondamentale n'est pas que l'absence de réponse à un item mais la position du sujet sur le trait latent (le score vrai de la psychométrie). Ainsi, au lieu de rechercher une estimation ponctuelle du score du manager, le 360 pourrait aboutir à une estimation de la distribution, conditionnellement à la variance intra. Ainsi, bien qu'il s'agisse d'une distribution construite à partir des scores observés et que cela ne s'inscrive plus, à proprement parler, dans la théorie du score vrai, nous avons eu recours à l'analyse factorielle multi-niveaux sur les réponses des observateurs aux items. En effet, cette analyse pourrait permettre de construire pour chaque leader une distribution dont l'étendue serait fonction de la variabilité de ses observateurs. Aussi, dans la dernière partie de ce chapitre, nous avons appliqué l'analyse factorielle confirmatoire multi-niveaux sur nos données, pour mettre en lumière les spécificités des scores aux 360 et comment cette démarche statistique permet d'appréhender ceux-ci aux deux niveaux d'analyse.

Discussion générale

Cette thèse, porte sur le leadership et plus précisément sa mesure, et la méthode d'évaluation à 360°.

Comme nous l'avons montré, le leadership est un concept à multiples facettes, qui recouvre un large domaine de définitions et de conceptions théoriques. Presque chacune de ces définition a donné naissance à des outils de mesure particuliers.

En étudiant ces outils nous avons pu constater la fréquence à laquelle l'évaluation du leadership nécessite l'intervention d'un tiers. En effet, pour la plupart des théories, la qualité de leader ne peut être pleinement appréciée qu'à travers la perception des subordonnés. C'est à ce moment que s'est posé la question de l'hétéro-évaluation, de sa fiabilité et des limites qui pouvaient y être associées.

En effet, l'hétéro-évaluation et surtout l'approche en 360° (évaluation d'un manager par un ensemble d'observateurs ayant avec lui des rapports hiérarchiques différents : subordonné, supérieur, collègue) est très largement utilisée pour l'évaluation du leadership. Les tenants de cette méthode apprécient tout particulièrement la multiplicité des points de vue que celle-ci permet. Ainsi, en plus d'une auto-évaluation effectuée par le manager lui-même, l'approche à 360° permet de recueillir des évaluations, sur les mêmes dimensions, réalisées par un ensemble d'observateurs. L'objectif premier de cette démarche est de confronter l'auto-évaluation du manager avec la perception qu'ont les personnes de son entourage professionnel de ses compétences managériales. Utilisé principalement dans le cadre de formations ou de sessions de développement professionnel, cette mise en perspective vise à aider le manager à prendre conscience des disparités pouvant exister entre sa propre perception et celles de ses observateurs, et à y faire face.

Un autre aspect est central aux démarches d'évaluation à 360° : la variété des catégories d'observateurs et donc la multiplicité des situations d'observation. Ainsi, le manager recevant le feed-back, a la possibilité de faire la part des choses entre les perceptions de ses subordonnés et celles de son supérieur et de ses collègues, qui – d'après les tenants de l'évaluation à 360° - présentent d'importantes différences de points de vue dues à des contextes et occasions d'observation différentes.

Dans le cadre de cette thèse, nous nous sommes particulièrement intéressés aux différences existant entre les évaluateurs d'un même manager. Différences que nous avons d'abord cherché à expliquer, comme l'ont fait de nombreux auteurs, par l'effet d'un ensemble de variables telles que la catégorie d'observateur, l'âge ou le genre. Ainsi, lorsque nous avons débuté notre travail sur les résultats des 360, notre premier intérêt a été de vérifier dans quelle mesure les variables évoquées pour expliquer les sur- et sous-évaluations de certains observateurs par rapport à d'autres, permettaient réellement de rendre compte des différences observées dans les scores. Et ce, en utilisant les méthodes statistiques les plus adaptées à l'analyse de ces données : les analyses multi-niveaux, méthodes qui ont été jusque là peu employées de manière aussi systématique que nous l'avons fait.

Aussi, en tenant compte de la structure particulière des données des 360, nous avons étudié l'effet sur les scores des observateurs de plusieurs variables classiquement utilisées pour commenter ou expliquer des différences d'évaluation : le genre, la position hiérarchique de l'observateur par rapport au self (catégorie) et l'âge.

Nos hypothèses construites à partir des effets identifiés dans la littérature ont été validées et nous avons observé des effets significatifs des caractéristiques des observateurs sur leurs jugements ou plutôt sur la variabilité des jugements. Pourtant, malgré la significativité de ces effets, les parts de variance que ces caractéristiques permettent d'expliquer ne dépassent jamais 2% de la variance totale – ce qui est peut être considéré comme descriptivement faible. Nous avons donc dû faire le constat qu'une grande partie de la variabilité au sein des observateurs d'un même self ne peut être expliquée par les caractéristiques les plus souvent invoquées et étudiées.

A partir de là, nous avons orienté notre recherche vers l'étude de cette variabilité et plus spécifiquement de l'accord et de la consistance entre les observateurs d'un même self. Car bien que partageant l'attrait du 360° et surtout de la multiplicité des évaluations que cette approche permet, nous avons été particulièrement intrigués par cette importance de variabilité restante entre les observateurs, même après avoir pris en compte des caractéristiques comme les catégories d'observateurs, leur genre et leurs âges.

En effet, comme nous l'avons rappelé, le 360° vise essentiellement à mettre le manager face aux écarts existant entre son auto-évaluation et l'hétéro-évaluation de ses observateurs. Pourtant, dans une majorité de cas, cette hétéro-évaluation est résumée par une moyenne arithmétique des réponses des différents observateurs. Or, nous avons montré que les valeurs des coefficients Intraclasse ne permettent pas de conclure à une consistance inter-juges robuste, il existe donc une forte variabilité (ne pouvant être expliquée par les caractéristiques des observateurs) entre les observateurs d'un même self. Dans une telle situation peut-on encore parler de consistance ou même simplement d'un consensus dans l'avis des observateurs ? En d'autres mots, au vu des faibles coefficients Intraclasse observés, peut-on réellement considérer que les observateurs sont en accord quand ils évaluent un même manager ? De ce fait l'utilisation d'un indice agrégé comme la moyenne se justifie-t-il ? est-il valide ?

Sur ces questions la littérature est quelque peu partagée. En effet, un grand nombre d'études ont conclu à un accord modéré entre les observateurs d'une même catégorie et à un accord faible entre les observateurs de catégories différentes. Cette conclusion était la position majoritaire jusqu'à ce que LeBreton et son équipe proposent une interprétation différente des valeurs communément obtenues.

D'après Le Breton, ces valeurs (faibles et modérées) ne seraient dues qu'à un artefact statistique lié à l'utilisation d'indices corrélationnels pour mesurer l'accord. Les auteurs ont donc proposé d'estimer l'accord avec un indice non corrélationnel basé sur une comparaison de la variance intra-groupe à la variance pouvant être obtenue par le hasard.

En appliquant l'approche de LeBreton nous avons, nous aussi, avec les indices r_{WG} , trouvé un niveau d'accord beaucoup plus élevé entre les observateurs comparativement aux indices corrélationnels (corrélation Bravais-Pearson, Corrélation Interclasse). Pourtant, des limites de cette approche nous sont également apparues. Nos conclusions ne nous ont donc pas permis, au final, de conclure avec conviction à l'existence d'un accord élevé entre les observateurs car trop de variance restait inexpliquée.

Il est à noter qu'aujourd'hui de plus en plus d'articles font références aux travaux de LeBreton et utilisent sa méthode pour apprécier l'accord entre les sources d'évaluation dans le cadre des 360. Nous avons donc choisi d'étudier plus en profondeur cette nouvelle

méthode, ce qui nous a permis de mettre en lumière un certain nombre de limites de celle-ci.

Ces conclusions seront utiles à ceux choisissant d'utiliser les indices r_{WG} pour apprécier l'accord entre évaluateurs, car elles nuancent les conclusions de LeBreton concernant l'accord observé. En effet, nous avons montré que le classement des groupes d'observateurs en tant que groupe en accord ou en désaccord, permet bien d'expliquer une part importante des écarts Inter-groupes (inter-self) mais ne permet pas d'expliquer une part notable des écarts intra-self (inter-observateur intra-self). De même, notre thèse a également mis en lumière l'existence d'une catégorie de groupes d'observateurs « Hors-Norme », présentant des caractéristiques spécifiques, les différenciant aussi bien des groupes d'observateurs en accord que des groupes d'observateurs en désaccord. Il serait donc utile dans les études faisant appel à l'approche de LeBreton, de porter attention à ces groupes et ne pas simplement les considérer comme en désaccord (en remettant ces coefficients à 0).

Cependant, toujours intéressés par le désaccord entre les observateurs, nous avons décidé d'approcher la question quelque peu différemment. Nous avons calculé, pour chacun des observateurs d'un groupe évaluant le même self, un écart absolu correspondant à l'écart absolu moyen de cet observateur avec tous les autres observateurs de son groupe, pris deux par deux et ce pour toutes les dimensions du GELI. Cette nouvelle variable nous a permis de saisir toute l'ampleur de la variabilité existante et de caractériser chaque observateur avec une valeur qui lui est propre.

Tout comme dans les analyses précédentes, nous avons utilisé les caractéristiques des observateurs et des managers évalués (âge, genre, relation hiérarchique) pour tenter d'expliquer la variance de cette nouvelle variable Écart. Or tout comme précédemment, l'ensemble de ces caractéristiques ne nous ont permis d'expliquer qu'une part très limitée des écarts observés (entre 1 et 2% de la variance totale de la variable écart est expliquée par ces variables).

De même, nous avons vérifié le pouvoir explicatif de la classification de LeBreton (les r_{WGS}) sur la variabilité existant entre les observateurs. Comme prévu, la classification selon les groupes Accord, Désaccord et Hors-Norme a permis d'expliquer une part importante

de la variance. Pourtant, les r_{WGS} étant calculés à partir de la variance moyenne de chacun des groupes d'observateurs comparé à une estimation de la variance théorique dans le cas de réponses aléatoires, nous avons montré que cette classification permet en fait d'expliquer une part très importante de la variance Inter-groupes, mais échoue à expliquer la variabilité Intra-groupe. Ainsi, si l'on tient compte de la classification Accord, Désaccord et Hors-norme issue des coefficients r_{WGS} , nous parvenons bien à catégoriser les groupes selon une tendance d'accord moyenne, mais nous échouons à réellement saisir le sens de ces écarts.

C'est face à ce constat que nous avons choisi de revenir aux fondements théoriques au cœur des évaluations à 360°, mais aussi de toutes les évaluations psychométriques : la théorie du score vrai et de l'erreur de mesure.

La théorie psychométrique classique définit clairement le score observé à un test psychométrique comme la somme d'un score vrai et d'une erreur de mesure. Pourtant la situation des scores des évaluations à 360° est plus complexe. Comme nous l'avons rappelé, l'objectif même des évaluations à 360° vise à mettre en lumière des différences entre plusieurs évaluations d'un même manager, la théorie classique se trouve donc inadaptée : soit il y a autant de scores vrais que d'observateurs, soit les écarts observés relèvent de l'erreur de mesure autour d'un seul score vrai.

Nous avons donc cherché à proposer une approche psychométrique du score au 360, non plus en tant qu'estimation ponctuelle du score vrai mais sous forme de distribution. Cette approche nous permet de considérer les scores des observateurs comme une distribution qui serait centrée autour du score vrai du manager, et dont la dispersion correspondrait à l'erreur.

Cette approche du score vrai sous forme de distribution, peut se modéliser grâce à l'analyse confirmatoire multi-niveaux. Celle-ci permet de distinguer le score vrai ainsi que l'erreur lié à l'Intra-groupe ainsi que le score vrai et l'erreur de l'Inter-groupe.

Cette conception théorique du score obtenu par les évaluations à 360° est un changement de perspective dans les évaluations multi-source. En effet, cette approche permet d'analyser et d'interpréter les divers effets à leur vrai niveau et d'approcher – de manière psychométrique – un outil qui manque de fondement quant à sa théorie métrique.

Ainsi, notre approche pourra permettre aux concepteurs de valider des questionnaires à 360° en estimant la vraie part d'erreur entachant leur mesure et donc en donnant des

indications claires quant à la fiabilité de leur instrument. De plus, notre thèse insiste sur l'intérêt, voire la nécessité, de considérer les scores au 360 en référence au modèle de la distribution, afin de s'approcher d'une estimation cohérente du score vrai de la personne évaluée. Enfin, notre travail met en lumière l'importance d'utiliser les méthodes statistiques adaptées aux résultats d'évaluations à 360°, que ce soit lors de l'analyse des effets (méthodes multi-niveaux), au moment de la validation (analyses confirmatoires multi-niveaux) mais aussi lors du calcul et de l'interprétation des scores.

- Les limites de notre étude

Une première limite pouvant être évoquée concernant notre étude porte sur les catégories d'observateurs. En effet, trois des quatre catégories d'observateur proposées, correspondent à des positions hiérarchiques spécifiques de l'observateur par rapport au self évalué : supérieur, subordonné ou collègue. C'est la quatrième catégorie, « Autre », qui pourrait-être mise en question. En effet, ce groupe est composé d'une grande variété de profils différents : clients, conjoints, fournisseurs... Le manque d'homogénéité au sein de cette catégorie rend difficile l'interprétation de ses effets. Qui plus est, cette variété entraîne également une variabilité accrue au sein du groupe.

Pourtant, comme nous l'avons montré, la variabilité est toute aussi présente dans les autres groupes d'observateurs qui sont plus homogènes quant à leur constitution. Qui plus est, dans un grand nombre de nos analyses, nous avons choisi d'étudier les observateurs et leurs réponses indépendamment de leur catégorie d'appartenance, la question du groupe « Autre » n'est donc apparue que de manière ponctuelle dans nos analyses.

En ce qui concerne l'outil que nous avons utilisé : le GELI, la passation en anglais peut, elle aussi, être considérée comme une limite. En effet, comme nous y reviendrons par la suite, notre échantillon est composé d'un grand nombre de nationalités différentes, dont une partie n'est pas de langue maternelle anglaise. Aussi, bien que la plupart des répondants travaillent en anglais ou en ont une bonne maîtrise, il n'est pas moins possible que ce facteur puisse être source de différences, notamment quant à la compréhension ou l'interprétation de certains items, créant ainsi des différences dans les résultats. En effet, lors des analyses factorielles confirmatoires, certains items présentent des liens modérés avec la dimension latente comparativement à d'autres, nous pourrions penser que ces

liens moindres puissent être dus à des différences de compréhension par exemple.

D'autre part, toujours concernant les observateurs, l'aspect interculturel peut être considéré comme un paramètre qui a été peu pris en compte dans notre étude.

En effet, le GELI s'adressant principalement aux *top-managers* exerçant dans des sociétés multinationales, il n'est pas rare que plusieurs nationalités soient représentées dans un groupe d'observateurs évaluant le même self. Ainsi, bien que la majorité des observateurs soient issus de pays occidentaux (Amérique du Nord, Europe, Australie) une partie non négligeable de l'échantillon provient, quant à elle, de pays très variés (Asie, Afrique, Moyen Orient...). De plus, au sein même des pays occidentaux, voire au sein même de l'Europe, certains peuvent mettre en avant l'existence de différences culturelles, nous pouvons donc penser que cette variété culturelle puisse avoir un impact sur les résultats et plus spécifiquement sur les différences observées dans les données.

Pourtant, malgré le large échantillon dont nous disposons, ces analyses n'ont pu être mises en place du fait de la trop faible occurrence des pays non occidentaux.

Cependant, dans une étude antérieure, il a pu être montré que dans l'ensemble, les effets identifiables par les différences entre nationalités étaient faibles (de l'ordre de 3%) (Vrignaud, 2008). Nous pouvons donc supposer que les effets de ces variables interculturelles ne permettraient pas, elles non plus, d'expliquer une part importante des variances observées.

De même, un élément contextuel pourrait expliquer en partie les scores globalement élevés observés dans les évaluations des managers. Les sels de notre échantillon sont tous des managers exerçant à des niveaux hiérarchiques très élevés et dans des sociétés de renom international, il n'est pas étonnant, dès lors, que ces managers présentent des évaluations et des scores les situant parmi les plus performants.

La question de la généralisabilité de nos résultats à d'autres questionnaires à 360° peut également se poser. En effet, un grand nombre d'outils existent pour évaluer le leadership et ce en 360°, nous pouvons donc nous demander si nos conclusions s'appliqueraient à d'autres instruments de mesure.

Tout d'abord il convient de rappeler que le GELI est un outil de mesure construit de manière classique, et bien que ses racines psychanalytiques le distinguent des autres outils d'évaluation du leadership, il s'inscrit clairement parmi les outils d'évaluation des

traits du leader. De même, le GELI présente d'excellentes qualités métriques ainsi qu'une structure robuste quant à ses échelles, il est donc de ce fait comparable à d'autres outils d'évaluation du leadership construits et validés selon les préceptes psychométriques. Enfin, au niveau descriptif, nous avons retrouvé dans les données du GELI un ensemble de caractéristiques largement présents dans la littérature pour les tests de leadership en 360 (effets des catégories d'observateur, utilisation des échelles de réponses...). Ces éléments nous font donc penser que nos conclusions seraient transférables à d'autres instruments de mesure du leadership mais aussi, certainement, aux 360 en général. Cela dit, il serait intéressant de vérifier nos conclusions sur des résultats issus d'un outil à 360° et portant sur une mesure plus large que le leadership – par exemple la personnalité. Nous pourrions ainsi généraliser nos résultats avec plus de certitude sur l'ensemble des outils à 360°.

D'autre part, à un niveau plus théorique, nous pouvons nous demander si finalement les différences observées dans les évaluations des observateurs sont de la nature des variables de la psychologie différentielle ou bien si elles sont plus des artefacts dus aux ancrages ou bien aux façons de noter des observateurs.

Il est intéressant de noter que la psychologie différentielle qui s'est longtemps centrée sur l'étude des différences Inter-individuelles notamment grâce aux tests, est aujourd'hui majoritairement investie dans l'étude de la variabilité Intra-individuelle.

De ce fait, l'étude des différences entre les observateurs prend un statut particulier car il s'agit bien de différences Inter-individuelles (Inter-observateur) mais en même temps, de différences Intra-individuelles (Intra-self), si l'on considère les évaluations des observateurs comme autant d'évaluations d'un même manager, on pourrait justement interpréter cette variabilité comme sa propre variabilité liée au contexte. Nous aurions donc ici une forme de variabilité intra- supportée par des variabilités inter-observateurs, justifiant d'autant plus le recours aux méthodes multi-niveaux pour l'analyse des résultats.

La question de la nature des différences dans les évaluations des observateurs reste pourtant posée et elle n'a pas trouvée pour nous de réponse définitive.

Ainsi, bien qu'une forte variabilité existe entre les observateurs d'un même self, la question peut se poser de savoir s'il s'agit de différences portant sur la dimension évaluée à proprement parler ou bien de différences liées aux stratégies de réponses.

Les tendances observées quant à l'utilisation de l'échelle, les biais divers pouvant être

associées avec les échelles de type Lickert (tendance à répondre aux extrêmes, tendance à répondre vers la moyenne, tendance à l'acquiescement...), les biais spécifiques aux 360° sont autant de stratégies – le plus souvent involontaires – pouvant être mises en œuvre par les répondants lors de leur évaluation.

Les différences observées pourraient donc refléter des stratégies différentes ou des biais différents plus que des différences quant au niveau du manager évalué sur la dimension latente.

Ce problème n'est pas récent et une grande partie de chercheurs ont fait remarquer les tendances fortement asymétriques dans les réponses aux évaluations à 360°. Pourtant peu de solutions ont été testées jusque là. Parmi celles-ci : l'incitation dans la consigne à utiliser l'ensemble des ancrages ou même l'utilisation dans le questionnaire d'ancrages uniquement positifs (allant, par exemples, pour une échelle de fréquence de parfois à toujours). Mais d'autres méthodes pourraient également être envisagées.

Nous avons proposé dans le chapitre 8, la technique des vignettes d'ancrage pour tenter de pallier ce problème, pourtant – malgré son intérêt – cette méthode risque de trouver ses limites dans le cadre du 360. En effet, cette technique visant à ramener les jugements sur une même échelle, risque de perdre son efficacité compte tenu du fait que l'étendue des jugements des observateurs est faible et que l'écart entre les vignettes devrait, lui, être maximal. Cela risque donc d'amener à un re-codage des jugements dans la catégorie haute pour une majorité d'observateurs. De la sorte, on risque d'aboutir à ce que les faibles, aient une meilleure variabilité – on retrouvera donc, dès lors, le même problème avec un faible nombre d'observateurs qui s'écartent de la majorité.

Les limites de ces techniques, nous ont donc conduit à changer de perspective quant au score du 360. Ne cherchant plus à obtenir une estimation ponctuelle du score vrai, nous avons choisi d'envisager ce score sous forme d'une distribution.

La question du score vrai se pose également dans la pratique, par exemple lors de l'étude de l'effet des formations. Les évaluations à 360° sont souvent réalisées en amont de sessions de formation ou de développement professionnel. Comme souvent, les formateurs et leurs commanditaires aiment évaluer l'effet de ces formations. Or la question du score se pose alors particulièrement. Que comparer ? Lorsque les évaluations pré et post formation ont toutes les deux été réalisées sous forme de 360°, le désaccord et la variabilité inhérente à ces évaluations va poser un sérieux problème si l'on cherche à calculer une estimation ponctuelle du score pré et post formation. Inversement, si l'on

envisage les scores obtenues lors de ces deux évaluations sous forme de distribution, alors, il deviendrait possible de comparer deux distributions pour chercher à conclure si une différence significative peut être observée.

Mais les questionnaires de leadership à 360° sont avant tout des mesures conatives, de personnalité. Aussi, certaines remarques propres aux tests de personnalité se posent également dans notre cas. La question classique du trait et de l'état concernant les dimensions de personnalité trouve sa place concernant la discussion sur l'évaluation du leadership. En effet, dans les évaluations psychométriques cette question se pose souvent lors de l'étude de la stabilité. Dans quelle mesure les changements observés lors de deux mesures consécutives sont-elles dues à un manque de fidélité de l'outil et dans quelle mesure à un manque de stabilité de la dimension mesurée ? La réponse la plus extrême à ce problème fut apportée par Spielberger (1983), dans son test STAI (*State Trait Anxiety Inventory*) où l'auteur propose deux versions du questionnaire une visant à mesurer un état d'anxiété général (Anxiété Trait), indépendant du contexte actuel et un autre visant à estimer un état d'anxiété situationnel (Anxiété État), fonction du moment de passation du questionnaire. Pourtant, contrairement à Spielberger, la majorité des chercheurs éludent la question situationnelle lors de la conception des tests. Le 360, cependant, remet au goût du jour ce débat de longue date. En effet, nous pouvons considérer les différentes hétéro-évaluations comme autant de contextes ou de situations particulières qui mériteraient d'être prises en compte.

Pourtant une autre question concernant le score dans le 360, en tant que mesure de personnalité, peut être posée : l'auto-évaluation est-elle de manière substantielle comparable à l'hétéro-évaluation ? Lors de l'auto-évaluation du manager, s'agit-il d'une estimation du score vrai sur le trait mesuré ou bien d'une estimation de la perception du manager de son niveau sur ce même trait ?

Cette question n'est pas spécifique au leadership et peut se poser pour toute évaluation de la personnalité, pourtant dans le 360, du fait des hétéro-évaluations cette question se pose particulièrement. Les évaluations des observateurs sont-elles comparables, en nature, avec l'auto-évaluation du manager ? Ou bien doit-on considérer ces deux évaluations séparément : les hétéro-évaluations comme des évaluations du niveau du manager sur le trait d'intérêt et l'auto-évaluation comme une évaluation de la perception

du manager de son niveau sur le trait d'intérêt ?

Le 360, en lui-même, ne permet pas plus qu'un autre outil d'auto-évaluation de répondre à cette question concernant les réponses du self. Pourtant cette différence pourrait justifier les comparaisons observateur-self, centrales dans les interprétations des résultats du 360. Ainsi, en considérant ces deux évaluations comme différentes en nature, nous pouvons envisager de les comparer et de nous intéresser aux différences – en tant que porteuses de sens.

Ainsi, le modèle de la distribution des scores semble être l'idéal psychométrique pour le 360, et permet si l'on considère que l'auto-évaluation et l'hétéro-évaluation différemment dans leur nature, à envisager un score de perception de soi, comparé à une distribution des scores des observateurs, centrées autour du score vrai du manager sur la dimension latente.

Au fond la question de la variabilité ne semble pas présenter de réponse univoque pour le 360. En effet, quelle serait au fond une bonne partition des variances entre la variance Inter- et Intra-groupe ? S'il n'y a que de la variance entre les selves et pas de variance entre les observateurs intra-self, alors quelle serait l'intérêt de mettre en place une procédure complexe et coûteuse visant à multiplier les observateurs ? Inversement, s'il y a trop de variance entre les observateurs au sein de chaque self, on risque d'obtenir des évaluations trop inconsistantes et manquant fondamentalement d'accord.

Il semble donc, que tel qu'utilisé aujourd'hui, le 360 doive présenter un compromis entre ces deux cas extrêmes : avoir de la variabilité, mais pas trop.

Index des Tableaux et Graphiques

Graphique T. 1 : Styles de leadership de Tannenbaum et Schmidt.....	30
Graphique T. 2 Styles de leader de Blake et Mouton.....	36
Graphique T.3 Modèle en 3D de Reddin.....	41
Graphique T. 4 Les Styles de leadership de Hershey et Blanchard.....	42
Tableau T. 1 Les liens entre les variables de la théorie « trajectoire-objectifs ».....	45
Tableau T. 2 Les styles de prise de décision de Vroom et Yetton	46
Tableau T.3 Caractéristiques personnelles, comportements, et effets du leader charismatique sur les subalternes.....	50
Graphique T.5 Théorie du Self-Concept de Shamir.....	53
Tableau W.1 : Comparaisons des dimensions de leadership transformationnel avec les dimensions d'intelligence émotionnelle selon 4 approches de ce concept.....	80
Tableau U.1 :Description des échelles du HDS en lien avec les troubles de personnalité du DSM.....	108
Tableau V.1 Fréquence des observations par sphères et par type d'observateur.....	128
Tableau A.1 Nombre d'items et coefficients alphas de Chronbach pour les 10 échelles du GELI.....	157
Tableau A. 2 Répartition de l'échantillon par type d'observateur.....	159
Tableau A. 3 Description de l'âge de l'échantillon.....	160
Tableau A. 4 Répartition de l'échantillon par genre selon le type d'observateur.....	160
Tableau A. 5 Répartition de l'échantillon par nationalité selon le type d'observateur.....	161
Tableau A. 6 Répartition de l'échantillon par Industries, selon le type d'observateur.	162
Graphique A.1 Diagramme d'une structure apparié	167
Tableau A. 7 Déviations (-2Loglikelihood) pour les modèles de niveau un et de niveau deux et les différences correspondantes pour les 10 échelles du GELI.....	172
Tableau A. 8 Variance et part de variance Intra et Inter pour les 10 échelles du GELI.....	174

Tableau A. 9 Réduction de la déviance (-2 Loglikelihood) entre le modèle nul et le modèle tenant compte des catégories d'observateur.....	177
Tableau A. 10 Coefficients de chacune des catégories d'observateurs pour les 10 échelles du GELI pour le modèle tenant compte de la catégorie d'observateur.....	178
Tableau A. 11 Pourcentage de variance totale expliquée par la variable Catégorie d'observateur pour chacune des 10 échelles du GELI.....	181
Tableau A. 12 Coefficients des variables âge du self et âge de l'observateur pour les 10 échelles du GELI pour le modèle tenant compte des âges.....	184
Tableau A. 12bis Réduction de la déviance (-2 Loglikelihood) entre le modèle nul et le modèle tenant compte des âges des observateurs et des selves, et pourcentage de variance expliqué par les variables âge.....	186
Tableau A. 13 Coefficients des catégories Homme manager évalué par une femme, femme manager évaluée par un homme, femme manager évaluée par une femme et différence entre les évaluations faites par les femmes observateur pour les managers hommes et femmes.....	190
Tableau A. 14 Coefficients des variables genre du self et genre de l'observateur pour les 10 échelles du GELI pour le modèle tenant compte des genres.....	192
Tableau A. 15 Réduction de la déviance (-2 Loglikelihood) entre le modèle nul et le modèle tenant compte du genre des observateurs et de celui des selves, et pourcentage de variance expliqué par les variables genre.	194
Tableau A.16 Réduction de la déviance (-2 Loglikelihood) entre le modèle nul et le modèle tenant compte des catégories d'observateur, du genre et de l'âge des observateurs et du genre et de l'âge des selves, et le pourcentage de variance expliqué par ces 5 variables.	196
Tableau A. 17 Pourcentage de la variance Intra, expliquée par les variables : catégorie d'observateur, âge et genre du self, âge et genre de l'observateur.....	198
Tableau B. 1 Corrélation moyenne, médiane et écart-type pour chaque paire d'observateur pour les 8 items de la dimension Visioning.....	203
Tableau B. 2 Corrélation moyenne et écart-type pour la dimension Visioning.....	204
Tableau B. 3 Corrélation moyenne, minimum, maximum et écart-type pour les 10	

dimensions du GELI.....	204
Tableau B. 4 Coefficient de corrélation Intraclasse pour chacun des items de la dimension Visioning.....	206
Tableau B. 5 Coefficient de Corrélation Intraclasse minimum et maximum sur les items de chacune des 10 échelles du GELI.....	206
Tableau B. 6 Coefficients $r_{wg(J)_{UN}}$, $r_{wg(J)_{SS}}$ et $r_{wg(J)_{MS}}$ pour chacune des 10 échelles du GELI.....	213
Tableau B. 7 Coefficient moyen et pourcentage d'observations présentant des coefficients supérieurs à 0,70 pour chacune des échelles du GELI et pour chacun des coefficients $r_{wg(J)_{UN}}$, $r_{wg(J)_{SS}}$, $r_{wg(J)_{MS}}$ et $r_{wg(J)_{HS}}$	216
Tableau B. 8 Variance d'erreur attendue (σ^2E) et proportion d'individus choisissant chacun des ancrages de réponse.....	218
Tableau B. 9 Distribution de la répartition des réponses dimension par dimension, pour les 7 ancrages disponibles pour les 17127 Observateurs.....	219
Tableau B. 10 Moyennes, médianes, skewness et kurtosis pour correspondant aux biais proposés par LeBreton et ceux observés pour notre échantillon pour les 10 dimensions.	221
Tableau B. 11 Moyenne, médiane et skewness attendus pour chacune des trois courbes asymétriques (légère, moyenne et forte) et pour pour la distribution des données expérimentales.....	222
Tableau B. 12 Pourcentages de groupes d'observateurs en Accord, en Désaccord ou Hors-Norme, pour chacune des 10 dimensions étudiées.....	224
Graphique B. 1 Pourcentage d'utilisation des 7 modalités de réponses aux 8 questions de la dimension Visioning par les observateurs, selon leur groupe « d'accord ».....	225
Tableau B. 13 Modalités des réponses présentant le plus de différences quant à leur fréquence d'utilisation, entre les groupes d'observateurs en Accord et ceux en Désaccord.	226
Tableau B. 14 Répartition des 2441 groupes d'observateurs selon leur propension d'accord.....	227
Tableau B. 15 Pourcentage de groupes ayant des Observateurs en catégorisés comme	

en « Accord » par les deux approches, pourcentage de groupes d'observateurs classés dans la même catégorie par les deux approches, Chi ² et le Tau-b pour chaque dimension.	228
Tableau B. 16 Coefficient de corrélation de Kendall pour chacune des dimensions prises deux par deux avec son seuil de significativité.....	230
Tableau C.1 Statistiques descriptives de la variables Écart pour la dimension Visioning.	234
Tableau C.2 Statistiques descriptives des 10 variables « Écart » pour les 10 dimensions du GELI.....	235
Tableau C.3 Statistiques description de la variables Écart Total.....	235
Tableau C. 4 Répartition de la variance en pourcentage entre l'Inter- et l'Intra-groupe pour les 10 dimensions du GELI.....	238
Tableau C.5 Coefficients pour les groupes d'accord (Désaccord et Hors-Norme) dans le modèle multi-niveaux tenant compte des groupes d'accord.....	239
Tableau C. 6 Pourcentage de la variance totale et de la variance inter-groupe expliqué par les catégories d'accord.....	241
Tableau C. 7 Réduction de la déviance (-2 Loglikelihood) entre le modèle nul et le modèle tenant compte des catégories d'observateur.....	244
Tableau C.8 Coefficients pour les catégories d'observateurs dans le modèle multi-niveaux tenant compte des catégories d'observateurs.....	245
Tableau C. 9 Réduction de la déviance (-2 Loglikelihood) entre le modèle nul et le modèle tenant compte des âges des observateurs et des selves.....	247
Tableau C. 10 Réduction de la déviance (-2 Loglikelihood) entre le modèle nul et le modèle tenant compte du genre des observateurs et des selves.....	248
Tableau C.11 Coefficients pour le genre dans le modèle multi-niveaux tenant compte du genre des selves et des observateurs.....	249
Tableau C. 12 Réduction de la déviance (-2loglikelihood) entre le modèle nul et le modèle avec les 5 variables explicatives et part de variance libre expliquée par les 5 variables explicatives ajoutées.....	252
Graphique Z.1 : Score aux vignettes d'ancrage et à un item du questionnaire de deux observateurs.....	267

Graphique Z. 2: Score re-codé des deux observateurs.	267
Graphique D.1 Résidus des évaluations faites par les observateurs pour la dimension Visioning.....	273
Tableau D.1 Coefficients de corrélation Intraclasse pour les items de la dimension Visioning.....	276
Graphique D. 2 Modèle théorique pour l'analyse factorielle confirmatoire multi-niveaux.	279
Tableau D.2 Indices CFI, RMSEA et SRMR pour la dimension Visioning.....	280
Graphique D. 3 Saturations du modèle factoriel confirmatoire multi-niveaux pour la dimension Visioning.....	281
Tableau D. 3 Parts de variance communes entre les items et les construits latents pour les 8 items de Visioning.....	282

Index des Équations

Équation A. 1 : Modèle d'une analyse de régression simple.....	158
Équation A. 2: Résidu.....	158
Équation A. 3 : Modèle d'une analyse à deux niveaux.....	159
Équation A. 4: Résidus du niveau groupe et du niveau individuel.....	159
Équation A. 5 Déviance du modèle à 1 niveau pour la dimension Visioning.....	162
Équation A. 6 Déviance du modèle à 2 niveaux pour la dimension Visioning.....	162
Équation A. 7 Réduction de la déviance entre le modèle à 1 et le modèle à 2 niveaux pour la dimension Visioning.....	162
Équation A. 8 Variance Inter- et Variance Intra-groupe pris en compte par le modèle pour la dimension Visioning.....	164
Équation A. 9 Variance totale estimée pour la dimension Visioning.....	164
Équation A. 10 Variance Partition Coefficient	164
Équation A. 11 VPC pour la dimension Visioning.....	165
Équation A. 12 Modèle pour la dimension Visioning tenant compte des catégories d'observateurs.....	167
Équation A. 13 Réduction de la déviance entre le modèle nul et le modèle tenant compte de la catégorie d'observateurs pour la dimension Visioning.....	167
Équation A. 14 Paramètres et variance inter- et intra-groupe du modèle tenant compte de la catégorie d'observateurs pour la dimension Visioning.....	168
Équation A. 15 Valeur des paramètres pour la catégorie Subordonné et Autres.....	168
Équation A. 16 Résidu de variance total après prise en compte de la variables Catégorie d'observateurs.....	170
Équation A. 17 Part de variance totale d'origine expliquée par l'ajout de la variable catégorie d'observateur.....	171
Équation A. 18 Modèle pour la dimension Visioning tenant compte de l'âge des observateurs et de l'âge du self.....	173
Équation A. 19 Réduction de la déviance entre le modèle nul et le modèle tenant compte	

des âges des observateurs et du self pour la dimension Visioning.....	174
Équation A. 20 Paramètres et variance inter- et intra-groupe du modèle tenant compte des âges des observateurs et du self pour la dimension Visioning.....	174
Équation A. 21 Valeur des paramètres pour les variables Âge des observateurs et Âge des selves.....	174
Équation A. 22 Résidu de variance total après prise en compte des variables âge des observateurs et âge du self.....	176
Équation A. 23 Part de variance totale d'origine expliquée par l'ajout des variables âge des observateurs et âge du self.....	176
Équation A. 24 Modèle pour la dimension Visioning tenant compte du genre des observateurs et du genre du self.....	179
Équation A. 25 Réduction de la déviance entre le modèle nul et le modèle tenant compte des genres des observateurs et du self pour la dimension Visioning.....	179
Équation A. 26 Paramètres et variance inter- et intra-groupe du modèle tenant compte des genres des observateurs et du self pour la dimension Visioning.....	179
Équation A. 27 Coefficient β_2	180
Équation A. 28 Différence entre les coefficient β_1 et β_3	180
Équation A. 29 Modèle pour la dimension Visioning tenant compte du genre des observateurs et du genre du self.....	182
Équation A. 30 Valeur des paramètres pour les variables Genre des observateurs et Genre des selves.	182
Équation A. 31 Résidu de variance total après prise en compte des variables genre des observateurs et genre du self.....	184
Équation A. 32 Part de variance totale d'origine expliquée par l'ajout des variables genre des observateurs et genre du self.....	184
Équation A. 33 Modèle général pour la dimension Visioning tenant compte de la catégorie d'observateur, des âges et des genres des observateurs et du genre du self.....	186
Équation A. 34 Réduction de la déviance entre le modèle nul et le modèle tenant compte de la catégorie d'observateur, des âges et des genres des observateurs et du genre du self pour la dimension Visioning.....	186

Équation A. 35 Paramètres et variance inter- et intra-groupe du modèle tenant compte de la catégorie d'observateur, des âges et des genres des observateurs et du genre du self pour la dimension Visioning.....	186
Équation A. 36 Résidu de variance total après prise en compte des 5 variables du modèle.	186
Équation A. 37 Part de variance totale d'origine expliquée par l'ajout des 5 variables explicatives.....	187
Équation A. 38 Part de variance Intra-groupe d'origine expliquée par l'ajout des 5 variables explicatives.....	188
Équation B. 1 Corrélation Bravais-Pearson.....	192
Équation B. 2 Corrélation Intraclasse One Way ANOVA.....	192
Équation B. 3 Corrélation Intraclasse sur des évaluations moyennes.....	193
Équation B. 4 Formule de r_{wg}	200
Équation B. 5 Formule de $r_{WG(J)}$	200
Équation B. 6 Formules des coefficients $r_{wg1 UN}$, $r_{wg1 SS}$, et $r_{wg1 MS}$	201
Équation B. 7 Formules des coefficients $r_{WG(J)_UN}$, $r_{WG(J)_SS}$ et $r_{WG(J)_MS}$	202
Équation C. 1 Modèle de niveau 1 et sa déviance pour la dimension Écart du facteur Visioning.....	225
Équation C. 2 Modèle de niveau 2 et sa déviance pour la dimension Écart du facteur Visioning.....	225
Équation C. 3 Variance Inter- et Variance Intra-groupe pris en compte par le modèle pour la dimension Écart du facteur Visioning.....	226
Équation C. 4 Variance totale estimée pour la dimension Écart du facteur Visioning.....	226
Équation C. 5 Variance Partition Coefficient	226
Équation C. 6 Paramètres pour les caractéristiques d'accord dans le modèle sur les écarts de la dimension Visioning.....	227
Équation C. 7 Valeurs des paramètres pour la catégorie Accord et Hors-norme.....	228
Équation C. 8 Variances inter- et intra-groupe du modèle tenant compte de la catégorie d'accord pour la dimension Écart du facteur Visioning.....	229

Équation C. 9 Variance Totale.	229
Équation C. 10 Part de la variance totale d'origine expliquée par les catégories d'accord.	229
Équation C. 11 Part de la variance Inter-self expliquée par le modèle.....	229
Équation C. 12 Modèle pour la dimension Écart du facteur Visioning tenant compte des catégories d'observateurs.....	231
Équation C. 13 Réduction de la déviance entre le modèle nul et le modèle de l'équation C.12.....	231
Équation C. 14 Paramètres associés au modèle de la dimension Écart du facteur Visioning.....	232
Équations C. 15 Valeurs des paramètres pour la catégorie Collègue et Subordonné.....	233
Équation C.15bis Modèle pour la dimension Écart du facteur Visioning tenant compte des âges des observateurs et des selves.....	234
Équation C.16 Réduction de la déviance entre le modèle nul et le modèle tenant compte des âges des observateurs et du self pour la dimension la dimension Écart du Visioning.	234
Équation C. 17 Modèle pour la dimension Écart du facteur Visioning tenant compte du genre des observateurs et du genre du self.	236
Équation C. 18 Réduction de la déviance entre le modèle nul et le modèle tenant compte des genres des observateurs et du self pour la dimension Écart du facteur Visioning.....	236
Équation C. 19 Paramètres du modèle tenant compte des genres des observateurs et du self pour la dimension Écart du facteur Visioning.....	237
Équation C. 20 Valeur des paramètres pour les variables Genre des observateurs et Genre des selves.	237
Équation C. 21 Modèle général pour la dimension Écart du facteur Visioning tenant compte de la catégorie d'observateur, des âges et des genres des observateurs et du genre du self.....	239
Équation C. 22 Réduction de la déviance entre le modèle nul et le modèle tenant compte de la catégorie d'observateur, des âges et des genres des observateurs et du genre du	

self.....	239
Équation C. 23 Part de variance totale d'origine expliquée par le modèle.....	239
Équation C. 24 Variance Inter- et Variance Intra-groupe du modèle pour la dimension Écart Total.....	241
Équation C. 25 Modèle général pour la dimension Écart Total tenant compte de la catégorie d'observateur, des âges et des genres des observateurs et du genre du self..	241
Équation C. 26 Valeur des paramètres pour les variables Collaborateur, Subordonné et Autre.....	242
Équation C. 27 Valeur des paramètres pour les variables Genre des observateurs et Genre des selves.....	242
Équation C. 28 Part de la variance d'origine expliquée par les cinq variables explicatives.	242
Équation C. 29 Variance Inter- et Variance Intra-groupe du modèle pour la dimension Écart Total.....	243
Équation C. 30 Valeur du paramètre pour la variable Nombre moyen d'utilisation de l'ancrage 2.....	243
Équation C. 31 Part de la variance inter et intra-self expliqué par la variable « Nombre d'utilisations moyennes de l'ancrage 2 ».....	244
Équation D. 1 Score vrai.....	247
Équation D. 2 Décomposition du score observé.....	249
Équation D. 3 L'erreur standard de mesure.....	249
Équation D. 4 Décomposition de la variance Totale.....	262
Équation D. 5 Corrélation.....	262
Équation D. 6 Corrélation Intraclasse.....	263

Index des Auteurs

Abrainis D. J.....	58, 301	Barbuto J. E.....	71, 303
Achenbach T. M.....	116, 301	Barchard K. A.....	71, 303
Adizes I.....	70, 301	Barkley R. A.....	115, 303
Ahn A.....	310	Barling J.....	72, 303
Aitkin M.....	157	Barrik M. R.....	65, 304
Ajzen I.....	77, 301	Bartram P.....	251
Alban-Metcalfe R. J.....	92, 93	Bass B. M...6, 48, 49, 51, 58, 60, 66, 67, 72,	
Aldag R. J.....	84, 306	89, 91, 112, 134, 178, 303, 304, 305, 307,	
Alimo-Metcalfe.....	92, 93, 301	311, 314, 317, 318, 319, 332, 334	
Allal L.....	140, 307	Batchelder C. S.....	256
Allen N. J.....	60, 62, 301, 307	Batchelder W. H.....	256
Alutto J.....	310	Beatty R. W.....	120, 121, 321
Amichai-Hamburger Y.....	71, 312	Bebb M.....	48, 305
Amoo T.....	251	Beehr T. A.....	134, 305
Anderson D.....	130, 131, 157	Behling O.....	16, 326
Anderson R. J.....	130, 131, 157, 301, 332	Bentler P. M.....	266
Antonakis J.....	67, 73, 301, 303, 328	Bernaud J. L.....	96, 305
Antonioni D.....	120, 302	Bernstein I. H.....	198, 261
Arad S.....	96, 302	BeShears R. S.....	71, 305
Arnold J.....	96, 302	Beyer J. M.....	40, 331
Ashford S. J.....	129, 302	Bharamanaikar S. R.....	71, 329
Ashkanasy N. M.....	63, 71, 302, 324	Biederman J.....	116, 305
Ashour A. S.....	30, 302	Blake R. R....5, 27, 28, 32, 81, 82, 282, 305	
Ashton M. C.....	68, 130, 302, 320, 332	Blanchard K. H.....	5, 34, 82, 83, 282, 315
Atwater L. E....66, 122, 123, 129, 131, 134,		Bland C.....	125, 311
135, 178, 302, 303, 304, 325, 333, 334		Block J.....	65, 128, 305, 327
Avery D. R.....	56	Bono J. E.....	56, 66, 68, 318
Avolio B. J....48, 58, 60, 62, 66, 67, 72, 88,		Borman W. C.....	119, 128, 139, 305, 306
89, 91		Boyatzis R.....	313
Ayman R.....	30, 303	Bradburn N. N.	251
Bailey C.....	126, 303	Brainerd C. J.....	116
Bar-On R.....	72, 304	Brief A. P.....	84, 306

Brovold H.....	70, 325	Coons A. E.....	16, 79, 314, 315
Brown F. W....	71, 88, 89, 130, 139, 249, 306, 324	Costa P. T.....	65, 66, 309, 313, 323
Brutus S.	127, 130, 131, 132, 136, 178, 183, 306	Craig S. B.....	141, 311
Bryant S. E.....	71, 306	Cripps B.....	120
Bryman A.	28, 29, 306	D'hoore W.....	56, 59
Burbach M. E.....	71, 303	Daft R. L.....	16, 17, 309
Burnham D. H.....	41, 323	Dansereau F.....	46, 59, 155, 310, 314, 334
Burns J. M.	6, 16, 46, 47, 48, 49, 76, 88, 307	Darnal J. S.....	56, 313
Bycio P.....	60, 61, 62, 307	Daubman K. A.....	130, 310
Caligiuri P. M.....	133, 307	Day D. V.....	133, 307
Cameron W. J..	15, 135, 160, 307, 326, 333	De Los Reyes A.....	115, 310
Campbell D. J.....	76, 127, 307	De Nisi A. S.....	119, 310
Cardinet J.....	140, 307	de Vries R. E....	12, 16, 52, 53, 69, 104, 105, 144, 146, 147, 148, 302, 310, 318, 319
Carless S. A.....	56, 57, 60, 91, 95, 96, 307	Demaree R. G.....	127, 138, 199, 317
Caruso D. R.....	72, 307, 323	Den Hartog D. N.....	91, 310
Cashman J. F.....	46, 84, 314	Devers K. J.....	58
Cattell R. B.....	19, 64, 69, 307, 308	Di Blas L.....	302
Cavazotte F.....	67, 68, 73, 308	Diefendorff J. M.....	141, 142, 310
Chan D.....	67, 137, 138, 308, 325	Dobbins G. H.....	122, 311
Chemers M. W.....	6, 31, 51, 52, 308, 312	Dodge G. E.....	155
Cheng B. S.....	122, 311	Dorfman P.....	135
Cheung G. W.....	141, 251, 308, 333	Doyle A.....	305
Cheung S.....	141, 251	Drasgow F.....	96, 141, 302, 310, 311
Christal R. E.....	64, 331	Dulewicz V.....	72, 311
Cleveland J. N.....	122, 131, 308, 324	Dyer N. G.....	269
Cole D. A.....	115, 308	Eber H. W.....	19, 308
Coleman M. M.....	131, 308	Edward J. R.....	124, 125, 311
Conger J. A....	5, 42, 43, 44, 71, 86, 87, 88, 112, 309	Edwards.....	137
Connelly M. S.....	56	Edwards J. E.....	137
Conway J.....	139	Edwards J.E.....	137
Cook M.....	115, 116, 120, 309, 314	Elliott S. N.....	115, 314
		Engle S.....	16, 327
		Erofeev D.....	134, 305

Evan A.....	124, 311	Griffin R. W.....	119, 310
Evans J. W.....	328	Griffith K. H.....	56
Facteau J. D.....	141, 311	Gudanowski D. M.....	134, 305
Faraone S. V.....	305	Gulliksen.....	247, 249
Farh J. L.....	122, 311	Gulliksen H.....	247
Feinberg B. J.....	131	Gupta V.....	135
Ferstl K. L.....	124, 318	Hackett R. D.....	60, 307
Fiedler F. E.....	5, 30, 31, 84, 85, 312	Haclonan J. R.....	59
Fiske D. W.....	76, 307	Haga W. J.....	46, 310
Fleenor J. W.....	122, 127, 183, 302, 306	Hall R. J... ..	34, 269, 310, 315, 332, 333, 334
Fleishman E. A.....	79, 80, 312	Halpin A. W.....	26, 79, 314
Fleiss J. L.....	138, 192, 193, 329	Hanges P.....	135, 269
Fletcher C.....	121, 126, 130, 303, 312, 318	Hanges P. G.....	135, 269, 316
Foa U. G.....	85, 312	Hannum K.....	132, 142, 313, 314
Fox S.....	71, 312	Hansen C. P.....	134, 305
Furman J. M.....	116, 312	Harris M. M.....	134, 314
Garman A. N.....	56, 57, 313	Harrison D. A.....	88, 306
Gasper S.....	91, 313	Hater J. J.....	58, 314
Gentry W. A.....	132, 313	Hattrup K.....	138, 139, 306, 319
George J. M.....	63, 73, 313	Hazucha J. F... ..	100, 101, 119, 123, 166, 315, 316, 327
Geyery A. L. J.....	56	Heatherington L.....	310
Gillies R. R.....	58	Hegarty W. H.....	123, 315
Goldberg L. R.....	65, 67, 68, 313	Heidemeier H.....	194
Goldstein M. J.....	116, 309	Hellevik L. W.....	119, 166, 315
Goleman D.....	72, 313	Hemphill J. K.....	16, 79, 315
Goodson J. R.....	84, 314	Hersey P.....	5, 34, 82, 83, 315
Gordon B.....	78	Heslin P. A.....	124, 316
Graeff C. L.....	83, 314	Hezlett S.....	100, 119, 315, 316
Graen G. B.....	46, 310, 314	Hickmann M.....	308
Graham J. W.....	259	Higgins C. A.....	42, 317
Greene C. N.....	29, 314	Higgs M.....	72, 311
Grégoire J.....	247	Hinde J.....	157
Greguras G. J.	141, 308, 310	Hoffman E.....	85, 316, 319
Gresham F. M.....	115, 314		

Hogan R.....	98, 99, 316, 323	112, 309	
Holt K. E.....	100	Katerberg R.....	29, 318
Hom P. W.....	29, 318	Katz D.....	24, 318
Horn J. L.....	141, 316	Kazdin A. E.....	115, 310
House R. J..	.5, 6, 16, 35, 36, 40, 41, 42, 44, 45, 73, 76, 135, 301, 316, 329	Kejnar M.....	61, 321
Howell J. M.....	42, 62, 88, 317	Keller T.....	67, 318
Hu L-T.....	266	Kelloway E. R.....	303
Huffcutt A.....	139, 309	Kenny D. A.....	127, 318
Hughes-James M.....	124, 323	Kerr R.....	71, 318
Humphrey R. H.....	71, 302	Kersting M.....	87, 88, 327
Hunt J. G.....	155	Kets de Vries M. F. R.....	12, 16, 52, 53, 104, 105, 144, 146, 147, 148, 318, 319
Huteau M.	115, 317	Kharti N.....	50, 51, 319
Ilgel D. R.....	129, 317	Khuntia R.....	56
Isaak J. L.....	131	King G.....	130, 252, 255
Ivanitskaya L.....	134, 305	Kingston P. W.....	132, 319
Jacob T.....	116, 312, 317	Klein L.....	138
Jacobs T. O.....	16, 317, 329	Knouse L. E.....	115, 303
Jacobsen E. N.....	83, 317	Kohlberg L.....	46, 47, 319
Jago A. G.....	5, 37, 38, 39, 332	Koopman P. L.....	91, 310
James L. R..	.6, 49, 124, 127, 138, 196, 199, 200, 317, 320, 323	Kouzes J. M.....	6, 49, 50, 93, 94, 319
Jaques E.....	317	Kozlowski S. W. G.....	138, 139, 306, 319
Javidan M.....	135	Kuhi L.....	125, 311
Jex S.....	130, 306	Lance C. E.....	141, 331
John O. P.....	65, 323, 327	Landy F. J.....	131, 308
Johnson J. W...	.34, 101, 124, 315, 318, 327	Larwood L.....	131, 320
Jones E. E.....	129, 130, 318, 323	Latham G. P.....	119, 124, 315, 316
Judge T. A.....	56, 66, 68, 318	Lautrey J.....	115, 317
Jung C. G.....	66, 69, 70, 303, 318	Laveault D.....	247
Jung D. I.....	66, 69, 70, 303, 318	Law K. S.....	63, 333
Jussim L.....	131	Leban W.....	71, 72, 320
Kahn R. L.....	24, 318	LeBreton J. M.....	9, 14, 128, 139, 143, 196, 198, 199, 200, 201, 203, 204, 205, 207, 209, 211, 213, 214, 221, 227, 228, 229,
Kanungo R. N..	5, 42, 43, 44, 71, 86, 87, 88,		

242, 245, 264, 273, 274, 284, 320	323
Lee K.....68, 127, 302, 307, 320	Mcclelland D. C.....41, 323
Lehman B. K.....305	McCrae R. R.....65, 66, 309, 313, 323
Lelouche C.....15, 320	McGee G. W.....84, 314
Lemoine C.....96, 305	McGraw K. O.....193
Levy P. E.....135, 320	McGregor D.....23, 76, 323
Lévy-Leboyer C.....103, 118, 320, 321	McKee A.....313
Lewin K.....5, 21, 22, 321	Meyer J. P.....62, 301
Likert R.....24, 25, 57, 111, 321	Miles R. H.....178, 251
Lim B. C.....67, 321, 325	Mills L. B.....71, 72, 73, 323, 333
Lindsey E.....103, 321	Misumi J.....56
Lippitt R.....321	Mitchell T. R.....85, 312
Loarer E.....251, 332	Montei M.....130, 306
Locke E. A.....59, 321	Moreno V.....308
Lodahl T. M.....61, 321	Morgeson F. P.....120, 121, 324
Loeber R.....116, 321	Moshavi D.....130, 324
Lombardo M.....101, 321	Mott P. E.....57, 324
London M.....120, 121, 123, 131, 134, 306, 307, 312, 321, 322, 328, 329, 331, 333	Möttus R.....255
Lord R. G.....63, 129, 322	Mount M. K.....65, 127, 304, 324
Lowe K. B.....91, 322	Mouton J. S.. 5, 27, 28, 32, 81, 82, 112, 282, 305
Luthans F.....125, 322	Mowday R. T.....62, 324
Mabe P. A.....128, 134, 135, 322	Murphy. 20, 26, 28, 115, 122, 170, 303, 307, 324, 328
Madden J. T.....77, 301	Murray C. J. L.....252
Mandell B.....322	Muthén B. O.....262, 263
Mann.....56, 63, 307, 322	Myers I. B.....70, 251, 324
Markham S. E.....155	Narayanan S.....84, 324
Marshall M.....56, 308	Newcombe M. J.....63, 318, 324
Marshall-Mies J. C.....56	Nielsen F.....131, 325
Maslow A. H.....24, 322	Nisbett R. E.....129, 318
Maurer T. J.....141, 142, 323	Nordvik H.....70, 325
Mayer J. D.....72, 307, 323	Norman W. X.....64, 65, 325
Mcardle J. J.....141, 316	Northouse P. G....17, 18, 37, 41, 42, 91, 325
McCauley C.. 101, 124, 127, 183, 306, 321,	

Nunally J. C.....	198, 261	Roman P. M.....	85, 316
O'Reilly C. A.....	131, 331	Romney A.....	256
Ohlott P.....	128, 331	Ronnkvist A. M.....	100
Oldham G. R.....	59	Rousseau D. M.....	58
Ones D. S.....	139	Rowold J.....	87, 88, 327
Ostroff C.....	131, 132, 173, 175, 325	Rubin D.....	10, 73, 145, 258, 259, 270, 327
Parry M. E.....	137, 311	Rushton J. P.....	116
Patterson C.....	91, 325	Sackett P. R.....	132, 327
Paulus C. J.....	16, 309	Sadler P. J.....	23, 328
Perugini M.....	302	Sala F.....	132, 328
Peter D. R.....	80, 307, 312	Salomon J. A.....	252
Peterson F. J.....	125, 322	Salovey P.....	72, 307, 323
Petty M. M.....	178	Sashkin M.....	50, 328
Pherwani S.....	72, 322	Schafer J. L.....	259
Ployhart R. E.....	67, 321, 325	Schaubroeck J.....	134, 314
Podsakoff P. M.....	42, 95, 325	Schenk J. A.....	59, 329
Ponnu C.....	89, 326	Schmidt F. L.....	22, 23, 139
Posner B. Z.....	6, 49, 50, 93, 94, 319	Schoeneman T. J.....	128, 329
Prati L. M.....	73, 326	Schriesheim C. A.....	26, 28, 46, 328
Pressley M.....	116	Schwarz J. C.....	115, 252, 328
Quinones M. A.....	56	Seeman M.....	135, 328
Radman H.....	124, 326	Self R. M....	6, 36, 44, 45, 91, 100, 137, 141, 145, 148, 150, 151, 152, 153, 154, 158, 175, 178, 181, 182, 183, 191, 200, 226, 237, 282, 302, 303, 305, 306, 307, 313, 315, 318, 320, 321, 328, 331, 332, 333, 334
Ragins B. R.....	57, 326	Senter J. L.....	139, 208, 320
Rasbash J.....	160	Sevy B. A.....	100, 328
Rauch C. F.....	16, 326	Shackleton V. J.....	29, 328
Reddin W. J.....	5, 32, 33, 34, 83, 282, 326	Shamir B.....	6, 44, 45, 73, 282, 329
Reilly.....	71, 125, 131, 306, 326, 329, 331	Shartle D.....	79
Rhoades J.....	96, 302	Shneider R.....	119, 315
Rice R. W.....	84, 326	Shortell S. M.....	58, 59, 60, 329
Richards D.....	16, 327	Shrauger J. S.....	128, 329
Riordan C. M.....	129, 327		
Rizzo J. R.....	6, 76, 316		
Roberson L.....	128, 327		
Robie C.....	101, 327		

Shrout P. E.....	138, 192, 193, 329	Trevino L. K.....	88
Silverman S. B.....	141, 310	Treviño L. K.....	88, 306
Simons T. L.....	58	Trice H. M.....	40, 331
Sivasubramaniam N.....	67, 91, 301, 322	Tse B.....	302
Skipton L. H.....	51, 329	Tsui A. S.....	128, 131, 178, 331
Slate F.....	303	Tupes E. C.....	64, 331
Smither J. W...122, 125, 134, 302, 322, 326,		Tyler J. L.....	56, 313
329, 333		Van Muijen J. J.....	91, 310
Son C.....	32, 68, 302	Vandenberg R. J.....	141, 331
Spangler W. D.....	40, 316	Vandenberghe C.....	56, 59, 60, 61, 62
Sparks J. R.....	59, 329	Vecchio R.....	85, 130, 131, 332
Spearman C.....	249, 251	Visser B. A.....	130, 332
Srivastava K. B.....	71, 329	Viswesvaran C.....	139, 332
Stalker G. M.....	76, 307	Vrignaud P.12, 148, 251, 259, 277, 319, 332	
Staw B. M.....	63, 302, 327, 330	Vroom V. H.....	5, 37, 38, 39, 282, 332
Sternberg R. J.....	303, 328	Walker A. G.....	125, 333
Stice G. F.....	19, 308	Wand J.....	252
Stogdill R. M.15, 16, 18, 26, 63, 80, 97, 314,		Wang M.....	122, 302
315, 330		Watkins D.....	251
Stordeur S.....	56, 59	Wearing A. J.....	56, 307
Suar D.....	56	Weaver G. R.....	88, 331
Szarota P.....	302	Weber M.....	40, 46, 86, 333
Taber T.....	78, 334	Weinstein N. D.....	128, 333
Tandon A.....	252	Weschler I. R.....	16, 330
Tannenbaum, R.....5, 16, 22, 23, 282, 330		West S. G.....	128, 134, 135, 322, 334
Tatsuoka M. M.....	19, 308	Westburgh E. M.....	30, 333
Tein J. Y.....	115, 116, 330	Wexley K. N.....	119, 315
Tejeda M. J.....	91, 330	Whetten D. A.....	135, 333
Tennakoon G.....	89, 326	White R. K.....	5, 21, 309, 321, 327
Thomas J. L.....	67, 305, 330	Winer B. J.....	26, 314
Tonidandel S.....	56, 139, 320	Wohlers A. J.....	131, 333
Tornow W. W.....	123, 124, 330, 331	Wolf G.....	127, 138, 199, 317
Tourneur Y.....	140	Wong C.-S.....	63, 193, 333
Tracey J. B.....	136, 334	Xie J. L.....	122, 333

Y. Tourneur Y.....	140, 307	Yip J.....	132, 313
Yammarino F. J..	58, 66, 123, 129, 134, 155, 161, 301, 303, 305, 310, 332, 334	Youtz M. A.....	129, 317
Yetton P. W.....	5, 37, 38, 282, 332	Yukl G.....	6, 17, 41, 77, 78, 97, 136, 334
Yielding D.....	251	Zagorsek H.....	94, 334
Yik M. S. M.....	122, 334	Zulauf C.....	71, 72, 320
		Waldman D. A....	48, 58, 123, 305, 332, 333

Références

- Abrainis, D. J. (1985). *Job Stressors, Strain, Job Performance, Social Support and Social Conflict: Causal Relationship in a Four-wave Longitudinal Panel Study* (Unpublished doctoral dissertation). University of Michigan - Cité par Shortel et al., 1991.
- Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist / 4-18 and 1991 profile*. Burlington, VT: University of Vermont.
- Adizes, I. (1979). *How to solve management crisis*. Los Angeles, CA: MNOR Institute.
- Adizes, I. (1987). *Corporate lifecycles*. Santa Monica, CA: Adizes Institute.
- Aitkin, M., Anderson, D., & Hinde, J. (1981). Statistical Modelling of Data on Teaching Styles. *Journal of the Royal Statistical Society. Series A (General)*, 144(4), 419-461.
- Ajzen, I., & Madden, T. J. (1986). Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. *Journal of Experimental Social Psychology*, 22(5), 453-474. doi:10.1016/0022-1031(86)90045-4
- Alimo-Metcalfe, B., & Alban-Metcalfe, R. J. (2001). The development of a new transformational leadership questionnaire. *Journal of Occupational and Organizational Psychology*, 74, 1 -27.
- Allen, N. J., & Meyer, J. P. (1990). The measurement and antecedents of affective, continuance and normative commitment to the organization. *Journal of occupational psychology*, 63(1), 1–18.
- Antonakis, J., Avolio, B. J., & Sivasubramaniam, N. (2003). Context and leadership: an examination of the nine-factor full-range leadership theory using the Multifactor Leadership Questionnaire. *The Leadership Quarterly*, 14(3), 261-295. doi:10.1016/S1048-9843(03)00030-4
- Antonakis, J., & House, R. J. (2002). An analysis of the full-range leadership theory: The way forward. In B. J. Avolio & F. J. Yammarino (Éd.), *Transformational and charismatic Leadership: The road ahead*, (p. 3-33). Amsterdam: Elsevier

Science/JAI.

- Antonioni, D. (1996). Designing an effective 360-degree appraisal feedback process. *Organizational Dynamics*, 25(2), 24-38. doi:10.1016/S0090-2616(96)90023-6
- Arnold, J., Arad, S., Rhoades, J., & Drasgow, F. (2000). The empowering leadership questionnaire: the construction and validation of a new scale for measuring leader behaviors. *Journal of Organizational Behavior*, 21, 249-269.
- Ashford, S. J. (1989). Self-assessments in organizations: A literature review and integrative model. In L. L. Cummings & B. W. Staw (Éd.), *Research in organizational behavior* (Vol. 11, p. 133–174). Greenwich, CT: JAI Press.
- Ashforth, B. E., & Humphrey, R. H. (1995). Emotion in the workplace: A reappraisal. *Human Relations*, 48, 97–126.
- Ashkanasy, N. M., & Tse, B. (2000). Transformational leadership as management of emotion: A conceptual review. In N. M. Ashkanasy, C. E. J. Härtel, & W. J. Zerbe (Éd.), *Emotions in the workplace: Research, theory, and practice* (p. 221–235). Westport, CT: Quorum Books.
- Ashour, A. S. (1973). The Contingency Model of Leadership Effectiveness: An Evaluation. *Organizational Behavior and Human Decision Processes*, 9(3), 339–355.
- Ashton, M. C., Lee, K., Perugini, M., Szarota, P., de Vries, R. E., Di Blas, L., Boies, K., et al. (2004). A six-factor structure of personality-descriptive adjectives: solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology*, 86(2), 356–366.
- Ashton, M. C., Lee, K., & Son, C. (2000). Honesty as the sixth factor of personality: correlations with machiavellianism, primary psychopathy, and social adroitness. *European Journal of Personality*, 14(4), 359–368.
- Atwater, L. E., Wang, M., Smither, J. W., & Fleenor, J. W. (2009). Are cultural characteristics associated with the relationship between self and others' ratings of leadership? *Journal of Applied Psychology; Journal of Applied Psychology*, 94(4), 876.

- Atwater, L. E., & Yammarino, F. J. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership predictions? *Personal Psychology*, 45, 141-164.
- Atwater, L. E., & Yammarino, F. J. (1993). Personal Attributes as Predictors of Superiors' and Subordinates' Perceptions of Military Academy Leadership. *Human Relations*, 46(5), 645-668. doi:10.1177/001872679304600504
- Atwater, L. E., & Yammarino, F. J. (1997). Self-other rating agreement. In G. R. Ferris (Éd.), *Research in personnel and human resources management* (Vol. 15, p. 121-174). Greenwich, CT: JAI Press.
- Avolio, B. J., Bass, B. M., & Jung, D. I. (1995). *Multifactor Leadership Questionnaire technical report*. Redwood City, CA: Mind Garden.
- Ayman, R. (2004). Situational and contingency approaches to leadership. In J. Antonakis, A. T. Cianciolo, & R. J. Sternberg (Éd.), *The nature of leadership* (p. 148-170). Thousand Oaks, CA: Sage Publications.
- Bailey, C., & Fletcher, C. (2002). The impact of multiple source feedback on management development: Findings from a longitudinal study. *Journal of Organizational Behavior*, 23(7), 853-867.
- Barbuto, J. E., & Burbach, M. E. (2006). The emotional intelligence of transformational leaders: A field study of elected officials. *The Journal of Social Psychology*, 146(1), 51-64.
- Barchard, K. A. (2003). Does Emotional Intelligence Assist in the Prediction of Academic Success? *Educational and Psychological Measurement*, 63(5), 840-858. doi:10.1177/0013164403251333
- Barkley, R. A., Knouse, L. E., & Murphy, K. R. (2011). Correspondence and disparity in the self-and other ratings of current and childhood ADHD symptoms and impairment in adults with ADHD. *Psychological assessment*, 23(2), 437.
- Barling, J., Slate, F., & Kelloway, E. R. (2000). Transformational leadership and emotional intelligence: An exploratory study. *Leadership and Organization Development*

Journal, 21(3), 157-161.

- Bar-On, R. (1997). *Bar-On Emotional Quotient Inventory: Technical manual*. Toronto, ON: Multi-Health Systems.
- Bar-On, R. (2006). The Bar-On model of emotional-social intelligence (ESI). *Psicothema*, 18 (supp)(1), 13-25.
- Barrik, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis., 44(1), 1-26.
- Bartram, P., & Yielding, D. (1973). The Development of an Empirical Method of Selecting Phrases Used in Verbal Rating Scales: A Report on a Recent Experiment. *Journal of the Market Research Society*, 15(3), 151-156.
- Bass, B. M. (1985). *Leadership and performance beyond expectations*. New York: Free Press.
- Bass, B. M. (1990). From transactional to transformational leadership: Learning to share the vision. *Organizational Dynamics*, 18(3), 19-31. doi:10.1016/0090-2616(90)90061-S
- Bass, B. M. (1995). Theory of transformational leadership redux. *The Leadership Quarterly*, 6(4), 463-478. doi:10.1016/1048-9843(95)90021-7
- Bass, B. M. (1998). *Transformational leadership: Industry, military, and educational impact*. Mahwah, NJ: Erlbaum.
- Bass, B. M., & Avolio, B. J. (1990a). *The multifactor leadership questionnaire*. Palo Alto, CA: Consulting Psychologist Press.
- Bass, B. M., & Avolio, B. J. (1990b). *Transformational Leadership Development: Manual for the Multifactor Leadership Questionnaire*. Palo Alto: CA: Consulting Psychologists Press.
- Bass, B. M., Avolio, B. J., & Atwater, L. E. (1996). The transformational and transactional leadership of men and women. *Applied Psychology: An International Review*, 45, 5-34.

- Bass, B. M., Waldman, D. A., Avolio, B. J., & Bebb, M. (1987). Transformational leadership and the falling dominoes effect. *Group and Organization Studies*, 12, 73–87.
- Bass, B. M., & Yammarino, F. J. (1991). Congruence of Self and Others' Leadership Ratings of Naval Officers for Understanding Successful Performance. *Applied Psychology*, 40(4), 437-454. doi:10.1111/j.1464-0597.1991.tb01002.x
- Beehr, T. A., Ivanitskaya, L., Hansen, C. P., Erofeev, D., & Gudanowski, D. M. (2001). Evaluation of 360 Degree Feedback Ratings: Relationships with Each Other and with Performance and Selection Predictors. *Journal of Organizational Behavior*, 22(7), 775-788.
- Bernaud, J. L., & Lemoine, C. (2000). *Traité de psychologie du travail et des organisations*. Paris: Dunod.
- BeShears, R. S. (2005). *The ability of emotional intelligence to predict transformational leadership when personality, affect, and cognitive ability are controlled*. (Dissertation Abstracts International). Section B: The Sciences and Engineering 65(10-B) - Cité par DeRue et al., (in press).
- Biederman, J., Faraone, S. V., Doyle, A., Lehman, B. K., Kraus, I., Perrin, J., & Tsuang, M. T. (1993). Convergence of the Child Behavior Checklist with Structured Interview-based Psychiatric Diagnoses of ADHD Children with and without Comorbidity. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, 34(7), 1241-1251. doi:Article
- Blake, R. R., & Mouton, J. S. (1961). *Group dynamics - Key to decision making*,. Houston, TX.: Gulf Publishing.
- Blake, R. R., & Mouton, J. S. (1978). *The new managerial grid*. Houston, TX.: Gulf Publishing Co.
- Block, J. (1961). *The Q-sort method in personality assessment and psychiatric research*. Springfield, IL: Charles C. Thomas.
- Borman, W. C. (1974). The rating of individuals in organisations : an alternative approach. *Organisational Behaviour and Human Performance*, 12, 105-124.

- Borman, W. C. (1997). 360 ratings: An analysis of assumptions and a research agenda for evaluating their validity. *Human Resource Management Review*, 7(3), 299–315.
- Bradburn, N. N., & Miles, C. (1979). Vague Quantifiers. *Public Opinion Quarterly*, 43(1), 92-101.
- Brief, A. P., & Aldag, R. J. (1981). The « Self » in Work Organizations: A Conceptual Review. *Academy of Management Review*, 6(1), 75-88. doi:10.5465/AMR.1981.4288006
- Brown, F. W., Bryant, S. E., & Reilly, M. D. (2006). Does emotional intelligence-as measured by the EQI-influence transformational leadership and/or desirable outcomes? *Leadership and Organizational Development*, 27(5), 300-351.
- Brown, K. G., Kozlowski, S. W. G., & Hattrup, K. (1996). Theory, issues, and recommendations in conceptualizing agreement as a construct in organizational research: The search for consensus regarding consensus. *Kozlowski, S. W. G. & Klein K. J. (Chairs) The meaning and measurement of within-group agreement in multi-level research*. Présenté à Symposium conducted at the annual convention of the Academy of Management Association, Cincinnati, OH. - Cité par Kozlowski et al., 1996.
- Brown, M. E., Treviño, L. K., & Harrison, D. A. (2005). Ethical leadership: A social learning perspective for construct development and testing. *Organizational Behavior and Human Decision Processes*, 97(2), 117-134. doi:10.1016/j.obhdp.2005.03.002
- Brutus, S., Fleenor, J. W., & McCauley, C. D. (1999). Demographic and personality predictors of congruence in multisource ratings. *Journal of Management Development*, 18, 417-435.
- Brutus, S., Montei, M., Jex, S., King, L., & King, D. (1993). Sex role egalitarianism as a moderator of gender congruence bias in evaluation. *Sex Roles*, 29(11-12), 755-765. doi:10.1007/BF00289216
- Bryman, A. (1992). *Charisma and Leadership in Organizations*. London: Sage Publications.

- Burns, J. M. (1978). *Leadership*. New York: Harper & Row.
- Burns, T., & Stalker, G. M. (1961). *The management of innovation*. London: Tavistock.
- Bycio, P., Hackett, R. D., & Allen, J. S. (1995). Further assessments of Bass's (1985) conceptualization of transactional and transformational leadership. *Journal of Applied Psychology, 80*(4), 468-478.
- Caligiuri, P. M., & Day, D. V. (2000). Effects of self-monitoring on technical, contextual, and assignment-specific performance. *Group and Organization Management, 25*, 164-174.
- Cameron, W. J. (1936). The American Leader. *Vital Speeches of the day.*, 599-600.
- Campbell, D. J., & Lee, C. (1988). Self-Appraisal in Performance Evaluation: Development Versus Evaluation. *Academy of Management Review, 13*(2), 302-314. doi:Article
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81-105. doi:10.1037/h0046016
- Cardinet, J., & Tourneur, Y. (1985). *Assurer la mesure*. Berne: Peter Lang.
- Cardinet, J., Tourneur, Y., & Allal, L. (1976). The symmetry of generalizability theory: applications to educational measurement. *Journal of educational measurement, 13*, 119-135.
- Carless, S. A. (1998). Assessing the discriminant validity of transformational leader behaviour as measured by the MLQ1. *Journal of Occupational & Organizational Psychology, 71*(4), 353-358. doi:Article
- Carless, S. A., Wearing, A. J., & Mann, L. (2000). A Short Measure of Transformational Leadership. *Journal of Business and Psychology, 14*(3), 389-405.
- Caruso, D. R., Mayer, J. D., & Salovey, P. (2002). Emotional intelligence and emotional leadership. In R. E. Riggio, S. E. Murphy, & F. J. Pirozzolo (Éd.), *Multiple intelligences and leadership* (p. 64-74). Mahwah, NJ: Erlbaum.
- Cattell, R. B., Cattell, A. K., & Cattell, H. E. (1993). *Sixteen Personality Factor*

Questionnaire, Fifth Edition. Champaign, IL: Institute for Personality and Ability Testing.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *Handbook for the 16PF*. Champaign, IL: Institute for Personality and ability testing.

Cattell, R. B., Marshall, M., & Greguras, G. J. (1957). Personality and motivation: Structure and measurement. *Journal of Personality Disorders*, 19(1), 53-67.

Cattell, R. B., & Stice, G. F. (1954). Four formulae for selecting leaders on the basis of personality. *Social Science Reporter*, 3, 14.

Cavazotte, F., Moreno, V., & Hickmann, M. (2012). Effects of leader intelligence, personality and emotional intelligence on transformational leadership and managerial performance. *The Leadership Quarterly*, 23(3), 443-455. doi:10.1016/j.leaqua.2011.10.003

Center for Creative Leadership. (2004). *Benchmarks facilitators' manual*. Greensboro, NC: Center for Creative Leadership.

Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology*, 83(2), 234.

Chemers, M. W. (1997). *An integrative theory of leadership*. Mahwah, NJ: Erlbaum.

Cheung, G. W. (1999). Multifaceted conceptions of self-other ratings disagreement. *Personnel Psychology*, 52(1), 1-36. doi:Article

Cleveland, J. N., & Landy, F. J. (1987). Age perception of jobs: Convergence of two questionnaires. *Psychological Reports*, 60(3c), 1075-1081. doi:10.2466/pr0.1987.60.3c.1075

Cole, D. A., Maxwell, S. E., & Martin, J. M. (1997). Reflected self-appraisals: Strength and structure of the relation of teacher, peer, and parent ratings to children's self-perceived competencies. *Journal of Educational Psychology*, 89(1), 55.

Coleman, M. M., Jussim, L., & Isaak, J. L. (1991). Black students' reactions to feedback

conveyed by White and Black teachers. *Journal of applied social psychology*, 21, 460-481.

Conger, J. A. (1991). Inspiring others : the language of leadership. *Academy of management executive*, 5, 31-45.

Conger, J. A., & Kanungo, R. N. (1987). Toward a Behavioral Theory of Charismatic Leadership in Organizational Settings. *Academy of Management Review*, 12(4), 637-647. doi:10.5465/AMR.1987.4306715

Conger, J. A., & Kanungo, R. N. (1994). Charismatic leadership in organizations: perceived behavioral attributes and their measurement. *Journal of Organizational Behavior*, 15, 439-352.

Conger, J. A., & Kanungo, R. N. (1997). Measuring charisma: Dimensionality and validity of the Conger-Kanungo scale of charismatic leadership. *Canadian Journal of Administrative Sciences*, 14(290-310).

Conway, J., & Huffcutt, A. (1997). Psychometric proprieties of multi-source performance ratings : A meta-analysis of subordinate, supervisor, peer and self-ratings. *Human Performance*, 10, 331-360.

Cook, M., & Cripps, B. (2005). *Psychological Assessment in the Workplace: A Manager's Guide*. New York: Wiley Editions.

Cook, W. L., & Goldstein, M. J. (1993). Multiple Perspectives on Family Relationships: A Latent Variables Model. *Child Development*, 64(5), 1377-1388. doi:Article

Costa, P. T., & McCrae, R. R. (1985). *The NEO Personality Inventory manual*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: PAR.

Daft, R. L. (2005). *The Leadership Experience* (3^e éd.). Mason, Ohio: Thompson South-Western.

Daft, R. L., & Paulus, C. J. (1994). *Making common sense: Leadership as meaning-*

making in a community of practice. Greensboro, NC: Center for Creative Leadership.

Dansereau, F., Alutto, J., & Yammarino, F. J. (1984). *Theory Testing in Organizational Behavior: The Variet Approach*. Englewood Cliffs, NJ: Perntice-Hall.

Dansereau, F., Graen, G. B., & Haga, W. J. (1974). A vertical dyad linkage approach to leadership within formal organizations: A longitudinal investigation of the role making process. *Organizational Behavior and Human Performance*, 13, 46-72.

Daubman, K. A., Heatherington, L., & Ahn, A. (1992). Gender and the self-presentation of academic achievement. *Sex Roles*, 27(3-4), 187-204. doi:10.1007/BF00290017

De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: a critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin; Psychological Bulletin*, 131(4), 483.

De Nisi, A. S., & Griffin, R. W. (2001). *Human resource management*. Boston: MA: Houghton Mifflin.

de Vries, R. E. (2012). Personality predictors of leadership styles and the self–other agreement problem. *The Leadership Quarterly*. doi:10.1016/j.leaqua.2012.03.002

Den Hartog, D. N., Van Muijen, J. J., & Koopman, P. L. (1997). Transactional versus transformational leadership: An analysis of the MLQ. *Journal of Occupational & Organizational Psychology*, 70(1), 19-34. doi:Article

Diefendorff, J. M., Silverman, S. B., & Greguras, G. J. (2005). Measurement Equivalence and Multisource Ratings for Non-managerial Positions: Recommendations for Research and Practice. *Journal of Business and Psychology*, 19(3), 399-425. doi:10.1007/s10869-004-2235-x

Drasgow, F. (1984). Scrutinizing psychological tests: Measurement equivalence and equivalent relations with external variables are the central issues.

Drasgow, F. (1987). Study of the measurement bias of two standardized psychological tests. *Journal of Applied Psychology*, 72(1), 19-29. doi:10.1037/0021-9010.72.1.19

- Dulewicz, V., & Higgs, M. (2000). Emotional intelligence a review and evaluation study. *Journal of Managerial Psychology, 15*(4), 341-372.
- Edward, M., Bland, C., & Kuhi, L. (1994). Academic administrator evaluating through colleague feedback. *CUPA Journal, 19*-29.
- Edward, M., & Evan, A. (1996). *360° feedback. The powerful new model for employee asesment and performance mprovement*. New York: Amacom.
- Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior and Human Processes, 58*, 51-100.
- Edwards, J. R. (1995). Alternatives to difference scores as dependent variables in the study of congruence in organizational research. *Organizational Behavior and Human Decision Processes, 64*, 307-324.
- Edwards, J. R. (2002). Alternatives to difference scores: Polynomial regression analysis and response surface methodology. In F. Drasgow & N. Schmitt (Éd.), *Measuring and analyzing behavior in organizations: Advances in measurement and data analysis* (p. 350-400). San Francisco, CA: Jossey-Bass, Inc.
- Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal, 36*(6), 1577-1613. doi:10.2307/256822
- ENOP. (1998). European Curriculum in W&O Psychology. Reference Model and Minimal Standards. Paris: Maison des Sciences de l'Homme.
- Facteau, J. D., & Craig, S. B. (2001). Are performance appraisal ratings from different rating sources comparable? *Journal of Applied Psychology, 86*(2), 215.
- Farh, J. L., & Cheng, B. S. (1997). Modesty bias in self-rating in Taiwan: Impact of item wording, modesty value, and self-esteem. *Chinese Journal of Psychology, 30*, 103-118.
- Farh, J. L., Dobbins, G. H., & Cheng, B. S. (1991). Cultural relativity in action: A comparison of self-rating made by Chinese and U.S. workers. *Personnel*

Psychology, 44(1), 129-147. doi:Article

Fiedler, F. E. (1967). *A Theory of Leadership Effectiveness*. New York: McGraw-Hill.

Fiedler, F. E. (1978). The Contingency Model and the Dynamics of the Leadership Process. *Advances in Experimental Social Psychology*, 11, 59-112.

Fiedler, F. E., & Chemers, M. W. (1984). *Improving leadership effectiveness: The leader match concept*. (2nd Ed.). New York: Wiley Editions.

Fleishman, E. A. (1953a). The description of supervisory behavior. *Journal of Applied Psychology*, 37(1), 1-6. doi:10.1037/h0056314

Fleishman, E. A. (1953b). The measurement of leadership attitudes in industry. *Journal of Applied Psychology*, 37(3), 153-158. doi:10.1037/h0063436

Fleishman, E. A. (1960). *Manual for Administering the Leadership Opinion Questionnaire*. Chicago: Science Research Associates.

Fleishman, E. A., & Peter, D. R. (1962). Interpersonal Values, Leadership Attitudes, and Managerial 'Success. *Personnel Psychology*, 15, 127-143.

Fletcher, C. (1997). *Appraisal: Routes to Improved Performance*. (2nd Ed.). London: IPD.

Fletcher, C. (1999). The implications of research on gender differences in self-assessment and 360 degree appraisal. *Human Resource Management Journal*, 9(1), 39-46. doi:10.1111/j.1748-8583.1999.tb00187.x

Foa, U. G., Mitchell, T. R., & Fiedler, F. E. (1971). Differentiation matching. *Behavioural Science*, 16, 130-142.

Fox, S., & Amichai-Hamburger, Y. (2001). The power of emotional appeals in promoting organizational change programs. *Academy of Management Executive*, 15(4), 84-95.

Friedman, H. H., & Amoo, T. (1999). Rating the rating scale. *Journal of Marketing Management*, 9(3), 114-123. doi:Article

Furman, J. M., & Jacob, R. G. (1997). Psychiatric dizziness. *Neurology May 1997*, 48(5), 1161-1166.

- Garman, A. N., Tyler, J. L., & Darnal, J. S. (2004). Development and Validation of a 360-degree-feedback Instrument for Healthcare Administrators. *Journal Of Healthcare Management, 49*(5), 307-322.
- Gasper, S. (1992). *Transformational leadership: An integrative review of literature*. (Doctoral dissertation). Western Michigan University.
- Gentry, W. A., Yip, J., & Hannum, K. (2010). Self-Observer Rating Discrepancies of Managers in Asia: A study of derailment characteristics and behaviors in Southern and Confucian Asia. *International Journal of Selection and Assessment, 18*(3), 237-250. doi:10.1111/j.1468-2389.2010.00507.x
- George, J. M. (1995). Leader positive mood and group performance. The case of customer service. *Journal of Applied Social Psychology, 25*, 778–794.
- George, J. M. (2000). Emotions and Leadership: The Role of Emotional Intelligence. *Human Relations, 53*(8), 1027-1055. doi:10.1177/0018726700538001
- Goldberg, L. R. (1983). *The magical number five, plus or minus two: Some considerations on the dimensionality of personality descriptors*. Paper presented at a Research Seminar, Gerontology Research Center, Baltimore, MD - Cité par Costa P. T., McCrae R. R. 1985.
- Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*(1), 26-34. doi:10.1037/0003-066X.48.1.26
- Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. Defruyt, & F. Ostendorf (Éd.), *Personality Psychology in Europe* (Vol. 7, p. 7-28). Tilburg, The Netherlands: Tilburg University Press.
- Goldberg, L. R., & Wheeler, L. (1981). Language and individual differences: The search for universals in personality lexicons. *Review of personality and social psychology* (Vol. 2, p. 141–165). Beverly Hills: CA: Sage Publications.
- Goleman, D., Boyatzis, R., & McKee, A. (2002). *Primal leadership: Realizing the power of emotional intelligence*. Boston: MA: Harvard Business School Press.

- Goodson, J. R., McGee, G. W., & Cashman, J. F. (1989). Situational leadership theory: a test of leadership prescriptions. *Group and organizational studies*, 14, 446-461.
- Graeff, C. L. (1997). Evolution of situational leadership theory: A critical review. *The Leadership Quarterly*, 8(2), 153-170. doi:10.1016/S1048-9843(97)90014-X
- Graen, G. B., & Cashman, J. F. (1975). A role making model of leadership in formal organizations: A developmental approach. In J. G. Hunt & L. L. Larson (Éd.), *Leadership Frontiers* (p. 143-165). Kent : Ohio: Kent State University Press.
- Graen, G. B., Dansereau, F., & Minami, T. (1972). Dysfunctional leadership styles. *Organizational Behavior and Human Performance*, 7, 216-236.
- Greene, C. N. (1975). The Reciprocal Nature of Influence Between Leader and Subordinate. *Journal of Applied Psychology*, 60(2), 187-193. doi:Article
- Gresham, F. M., Elliott, S. N., Cook, C. R., Vance, M. J., & Kettler, R. (2010). Cross-informant agreement for ratings for social skill and problem behavior ratings: An investigation of the Social Skills Improvement System—Rating Scales. *Psychological assessment*, 22(1), 157.
- Halpin, A. W. (1957). *Manual for the Leader Behaviour Description Questionnaire*. Columbus, Ohio: Fisher College of Business The Ohio State University.
- Halpin, A. W., & Winer, B. J. (1957). A factorial study of the leader behavior descriptions. In R. M. Stogdill & A. E. Coons (Éd.), *Leader behavior: Its description and measurement*. Columbus, Ohio - Bureau of Business Research: Ohio State University.
- Hannum, K. (2007). Measurement Equivalence of 360°-Assessment Data: Are different raters rating the same constructs? *International Journal of Selection and Assessment*, 15(3), 293–301.
- Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer and peer-supervisor ratings. *Personnel Psychology*, 41(1), 43-62. doi:Book Review
- Hater, J. J., & Bass, B. M. (1988). Superiors' evaluations and subordinates' perceptions of transformational and transactional leadership. *Journal of Applied psychology*, 73(4),

695.

- Hazucha, J. F., Hezlett, S. A., & Shneider, R. J. (1993). The impact of 360-degree feedback on management skills development. *Human Resource Management, 32*, 325-351.
- Hazucha, J. F., Latham, G. P., & Wexley, K. N. (1982). *Increasing productivity through performance appraisal*. Reading, MA: Addison-Wesley.
- Hegarty, W. H. (1974). Using subordinate ratings to elicit behavioral changes in supervisors. *Journal of Applied Psychology, 59*(6), 764.
- Heidemeier, H., & Moser, K. (2009). Self–other agreement in job performance ratings: A meta-analytic test of a process model. *Journal of Applied Psychology, 94*(2), 353–370.
- Hellevik, L. W., Hazucha, J. F., & Shneider, R. J. (1992). Behaviour change: Models, methods and a review of evidence. In M. D. Dunnette & L. Hough (Éd.), *Handbook of industrial and organisational psychology* (Vol. 3, p. 823-893). Palo Alto, CA: Consulting Psychologist Press.
- Hemphill, J. K., & Coons, A. E. (1957). Development of the Leader Behavior Description Questionnaire. In R. M. Stogdill & A. E. Coons (Éd.), *Leader Behavior: Its Description and Measurement* (Monograph No. 88.). Columbus, Ohio State University: Bureau of Business Research - Cité par Stogdill, R. M. 1969.
- Hersey, P., & Blanchard, K. H. (1969). Life cycle theory of leadership: Is there a best style of leadership? *Training and Development Journal, 33*(6), 26-34.
- Hersey, P., & Blanchard, K. H. (1974). So you want to know your leadership style? *Training and development journal, Feb.*, 22-37.
- Hersey, P., & Blanchard, K. H. (1982). *Management of organizational behavior: Utilizing human resources*. (4th ed.). Englewood Cliffs, NJ: Prentice Hall.
- Hersey, P., Blanchard, K. H., & Johnson, D. E. (2007). *Management of Organizational Behavior: Leading human resources*. (8th ed.). Englewood Cliffs, NJ: Prentice Hall.

- Heslin, P. A., & Latham, G. P. (2004). The effect of upward feedback on managerial behavior. *Applied Psychology, 53*(1), 23–37.
- Hezlett, S. H., Ronnkvist, A. M., Holt, K. E., & Hazucha, J. F. (1997). *The PROFILOR(R) technical summary*. Minneapolis, MN: Personnel Decisions International.
- Hoffman, E., & Roman, P. M. (1984). Criterion related validity of the least preferred co-worker measure. *The journal of social psychology, 122*, 79-84.
- Hogan, R., Curphy, G. J., & Hogan, J. (1994). What we know about leadership: Effectiveness and personality. *American Psychologist, 49*(6), 493-504. doi:10.1037/0003-066X.49.6.493
- Hogan, R., & Hogan, J. (1997). *Hogan Development Survey Manual*. Tulsa, OK: Hogan Assessment Systems.
- Hogan, R., & Hogan, J. (2001). Assessing Leadership: A View from the Dark Side. *International Journal of Selection and Assessment, 9*(1&2), 40-51. doi:10.1111/1468-2389.00162
- Horn, J. L., & Mcardle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*(3), 117-144. doi:10.1080/03610739208253916
- House, R. J. (1971). A Path Goal Theory of Leader Effectiveness. *Administrative Science Quarterly, 16*(3), 321-339. doi:Article
- House, R. J. (1977). A 1976 theory of charismatic leadership. In J. G. Hunt & L. L. Larson (Éd.), *Leadership: the cutting edge* (p. 189-207). Carbondale, IL: Southern Illinois University Press.
- House, R. J., Hanges, P., Javidan, M., Dorfman, P., & Gupta, V. (2004). *Culture, leadership, and organizations*. Thousand Oaks, CA: Sage Publications.
- House, R. J., & Rizzo, J. R. (1972). Towards the measurement of organizational practices: scale development and validation. *Journal of Applied Psychology, 56*(5), 388-396.
- House, R. J., Spangler, W. D., & Woycke, J. (1991). Personality and charisma in the U.S.

presidency: A psychological theory of leadership effectiveness. *Administrative Science Quarterly*, 36, 364-396.

Howell, J. M., & Avolio, B. J. (1992). The ethics of charismatic leadership: submission or liberation? *Executive*, 6(2), 43-54. doi:10.5465/AME.1992.4274395

Howell, J. M., & Higgins, C. A. (1990). Leadership behaviors, influence tactics, and career experiences of champions of technological innovation. *The Leadership Quarterly*, 1(4), 249-264. doi:10.1016/1048-9843(90)90004-2

Hunt, J. G., & Dodge, G. E. (2000). Leadership déjà vu all over again. *The Leadership Quarterly*, 11(4), 435-458. doi:10.1016/S1048-9843(00)00058-8

Huteau, M., & Lautrey, J. (1999). *Approches différentielles en psychologie*. Rennes: PUR.

Ilgen, D. R., & Youtz, M. A. (1989). Factors affecting the evaluation and development of minorities in organizations. In K. M. Rowland & G. R. Ferris (Éd.), *Research in personnel and human resources management* (Vol. 4, p. 307-337). Greenwich, CT: JAI Press.

Jacob, T., Tennenbaum, D. L., & Krahn, G. (1987). Factors influencing the reliability and validity of observational data. In T. Jacob (Éd.), *Family interaction and psychopathology* (p. 297-328). New York: Plenum.

Jacobs, T. O., & Jaques, E. (1991). Executive leadership. In R. Gal & A. D. Manglesdorff (Éd.), *Handbook of military psychology* (p. 431-447). Chichester, England: Wiley Editions.

Jacobsen, E. N. (1984). *The subordinate: A moderating variable between leader behavior and effectiveness*. (Dissertation Abstracts International). 45(7B), 2296 - Cité par Bass B. M. et al., (2008).

James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating Within-Group Interrater Reliability With and Without Response Bias. *Journal of Applied Psychology*, 69(1), 85-98. doi:Article

James, L. R., Demaree, R. G., & Wolf, G. (1993). r-sub(wg): An assessment of within-group interrater agreement. *Journal of Applied Psychology*, 78(2), 306-309.

doi:10.1037/0021-9010.78.2.306

- Johnson, J. W., & Ferstl, K. L. (1999). The effects of interrater and self-other agreement on performance improvement following upward feedback. *Personnel Psychology, 52*(2), 271-303. doi:Article
- Jones, E. E., & Fletcher, C. (2002). Self-assessment in a selection situation: An evaluation of different measurement approaches. *Journal of Occupational and Organizational Psychology, 75*, 145-161.
- Jones, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. Kanouse, H. H. Kelly, R. E. Nisbett, S. Valins, & B. Weiner (Éd.), *Attribution: Perceiving the causes of behavior* (p. 79-94). Morristown, NJ: General Learning Press.
- Judge, T. A., & Bono, J. E. (2000). Five-Factor Model of Personality and Transformational Leadership. *Journal of Applied Psychology, 85*(5), 751-765. doi:Article
- Jung, C. G. (1921). *Types psychologiques*. Genève: Georg Editeur.
- Katerberg, R., & Hom, P. W. (1981). Effects of Within-Group and Between-Groups Variation in Leadership. *Journal of Applied Psychology, 66*(2), 218-223. doi:Article
- Katz, D., & Kahn, R. L. (1952). Some recent findings in human relations research. In E. Swanson, T. Newcombe, & E. Hartley (Éd.), *Readings in social psychology*. NY, Holt: Reinhart and Winston.
- Keller, T. (1999). Images of the familiar: Individual differences and implicit leadership theories. *The Leadership Quarterly, 10*(4), 589-607. doi:10.1016/S1048-9843(99)00033-8
- Kenny, D. A. (1991). A general model of consensus and accuracy in interpersonal perception. *Psychological Review, 98*(2), 155-163. doi:10.1037/0033-295X.98.2.155
- Kerr, R., Gavin, J., Heaton, N., & Boyle, E. (2005). Emotional intelligence and leadership effectiveness. *Leadership & Organization Development Journal, 27*(2), 265-279.
- Kets de Vries, M. F. R. (1993). *Leaders, Fools, and Impostors*. San Francisco, CA: Jossey-

Bass.

- Kets de Vries, M. F. R. (2003a). *Dysfunctional Leadership*. (Working paper series No. 2003/58/ENT). INSEAD.
- Kets de Vries, M. F. R. (2003b, février 1). Etre un bon chef: essai de définition. *L'expansion*.
- Kets de Vries, M. F. R., Korotov, K., & Florent-Treacy, E. (2007). *Coach and Couch: The Psychology of Making Better Leaders*. Hants, UK: Palgrave Macmillan.
- Kets de Vries, M. F. R., Vriegnaud, P., & Florent-Treacy, E. (2004). The Global Leadership Life Inventory: development and psychometric properties of a 360-degree feedback instrument. *The International Journal of Human Resource Management*, 15(3), 475-492. doi:10.1080/0958519042000181214
- Kharti, N. (2005). An alternative model of transformational leadership. *Vision - The Journal of Business Perspective*, 9, 19-26.
- Kingston, P. W., Hubbard, R., Lapp, B., Schroeder, P., & Wilson, J. (2003). Why education matters. *Sociology of Education*, 76, 53-70.
- Kohlberg, L. (1964). Development of moral character and moral ideology. In M. L. Hoffman & L. W. Hoffman (Éd.), *Review of child development research* (Vol. 1, p. 383-432). New York: Sage Publications.
- Kouzes, J. M., & Posner, B. Z. (1987). *The leadership challenge: how to get extraordinary things done in organizations*. San Francisco, CA: Jossey-Bass.
- Kouzes, J. M., & Posner, B. Z. (1993). *Psychometric proprieties of the leadership practice inventory*. San Diego: Pfeifer & Company.
- Kozlowski, S. W. G., & Klein, L. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal and emergent processes. In K. J. Klein & S. W. G. Kozlowski (Éd.), *Multilevel theory, research, and methods in organizations* (p. 349-381). San Francisco, CA: Joesay-Bay.
- Kozlowski, S. W., & Hattrup, K. (1992). A disagreement about within-group agreement:

Disentangling issues of consistency versus consensus. *Journal of Applied Psychology*, 77(2), 161.

Larwood, L. (1982). The importance of being right when you think you are: Self-serving biases in equal employment opportunity. In B. A. Gutek (Éd.), *Sex role stereotyping and affirmative action policy* (p. 65-80). Los Angeles, CA: Institute of Industrial Relations.

Laveault, D., & Gregoire, J. (1998). *Introduction aux theories des tests en psychologie et en education*. (2nd Ed.). Bruxelles: De Boeck.

Leban, W., & Zulauf, C. (2004). Linking emotional intelligence abilities and transformational leadership styles. *The Leadership and Organization Development Journal*, 25(7), 554-564.

LeBreton, J. M., Burgess, J. R. D., Kaiser, R. B., Atchley, E. K., & James, L. R. (2003). The restriction of variance hypothesis and interrater reliability and agreement: Are ratings from multiple sources really dissimilar? *Organizational Research Methods*, 6, 78–126.

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, 11, 815-852.

LeBreton, J. M., & Tonidandel, S. (2008). Multivariate relative importance: Extending relative weight analysis to multivariate criterion spaces. *Journal of Applied Psychology*, 93(2), 329.

Lee, K., & Ashton, M. C. (2004). Psychometric Properties of the HEXACO Personality Inventory. *Multivariate Behavioral Research*, 39(2), 329-358. doi:10.1207/s15327906mbr3902_8

Lelouche, C. (2007). *Psychologie sociale des organisations*. Paris: Armand Colin.

Levy, P. E. (1993). Self-appraisal and attributions: A test of a model. *Journal of Management*, (19), 51-62.

Lévy-Leboyer, C. (2000). *Le 360° Outil de développement personnel*. Paris: Editions de l'organisation.

- Lévy-Leboyer, C. (2001). *Triptyque questionnaire 360° à modules pour les cadres dirigeants*. Paris: ECPA.
- Lewin, K., Lippitt, R., & White, R. K. (1939). Patterns of aggressive behavior in experimentally created social climate. *Journal of Social Psychology, 10*(2), 271-299.
- Likert, R. (1961). *New Patterns of Management*. New York: McGraw-Hill.
- Likert, R. (1967). *The human organization: Its management and value*. New York: McGraw-Hill.
- Lim, B. C., & Ployhart, R. E. (2003, novembre 3). *Transformational Leadership: Relations to the Five Factor Model and Team Performance in Typical and Maximum Contexts*. Présenté à The 45TH International military testing association conference,, FLORIDA PENSACOLA.
- Lindsey, E., Homes, V., & McCall, M. W. J. (1987). *Key events in executives' lives*. (No. Tech. Rep. No. 32). Greensboro, NC: Center for Creative Leadership.
- Locke, E. A. (1976). The nature and causes of job satisfaction. In M. D. Dunnette (Éd.), *Handbook of Industrial and Organizational Psychology* (p. 1297-1350). Chicago: Rand McNally.
- Lodahl, T. M., & Kejnar, M. (1965). The definition and measurement of job involvement. *Journal of applied psychology, 49*(1), 24.
- Loeber, R., Green, S. M., Lahey, B. B., & Stouthamer-Loeber, M. (1991). Differences and similarities between children, mothers, and teachers as informants on disruptive child behavior. *Journal of Abnormal Child Psychology, 19*(1), 75-95. doi:10.1007/BF00910566
- Lombardo, M., & McCauley, C. (1994). *Benchmarks: A manual and trainer's guide*. Greensboro, NC: Center for Creative Leadership.
- London, M. (1995). *Self and interpersonal insight: How people learn about themselves and others in organizations*. New York: Oxford University Press.
- London, M., & Beatty, R. W. (1993). 360-degree feedback as a competitive advantage.

Human Resource Management, 32, 352–373.

- London, M., & Smither, J. W. (1995). Can multi-source feedback change perceptions of goal accomplishment, self-evaluation, and performance-related outcome? Theory-based applications and directions for research. *Personnel Psychology*, 48(4), 803-839. doi:Article
- Lord, R. G., de Vader, C. L., & Alliger, G. M. (1986). A meta-analysis of the relation between personality traits and leadership perceptions: An application of validity generalization procedures. *Journal of Applied Psychology*, 71(3), 402-410. doi:10.1037/0021-9010.71.3.402
- Lowe, K. B., Kroeck, K. G., & Sivasubramaniam, N. (1996). Effectiveness correlates of transformational and transactional leadership: A meta-analytic review of the mlq literature. *The Leadership Quarterly*, 7(3), 385-425. doi:10.1016/S1048-9843(96)90027-2
- Luthans, F., & Peterson, F. J. (2003). 360-degree feedback with systematic coaching: empirical analysis suggests a winning combination. *Human Resource Management*, 42, 234-256.
- Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of applied Psychology*, 67(3), 280.
- Mandell, B., & Pherwani, S. (2003). Relationship Between Emotional Intelligence and Transformational Leadership Style: A Gender Comparison. *Jouranl Of Business and Psychology*, 17(3), 487-404.
- Mann, R. D. (1959). A review of the relationships between personality and performance in small groups. *Psychological Bulletin*, 56(4), 241-270. doi:10.1037/h0044587
- Markham, S. E. (2010). Leadership, levels of analysis, and déjà vu: Modest proposals for taxonomy and cladistics coupled with replication and visualization. *The Leadership Quarterly*, 21(6), 1121–1143.
- Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, 50(4), 370-396. doi:10.1037/h0054346

- Maurer, T. J., Raju, N. S., & Collins, W. C. (1998). Peer and subordinate performance appraisal measurement equivalence. *Journal of Applied Psychology, 83*(5), 693-702. doi:10.1037/0021-9010.83.5.693
- Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey & D. J. Sluyter (Éd.), *Emotional development and emotional intelligence* (p. 3-31). New York: Basic Books.
- Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion, 3*(1), 97-105. doi:10.1037/1528-3542.3.1.97
- McCauley, C. D., & Hughes-James, M. (1994). *An evaluation of outcomes of a leadership development program*. Greensboro, NC: Center for Creative Leadership.
- Mcclelland, D. C., & Burnham, D. H. (1976). Power is the great motivator. *Harward Business Review, 54*, 100-110.
- McCrae, R. R., & Costa, P. T. (1985). Openness to experience. In R. Hogan & W. H. Jones (Éd.), *Perspectives in personality* (Vol. 1, p. 145-172). Greenwich, CT: CXJAI Press.
- McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*(1), 81-90. doi:10.1037/0022-3514.52.1.81
- McCrae, R. R., Costa, P. T., & Busch, C. M. (1986). Evaluating comprehensiveness in personality systems: The California Q-Set and the five factor model. *Journal of Personality, 54*, 430-446.
- McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60*(2), 175-215.
- McGregor, D. (1960). *The Human Side of Enterprise*. New York: McGraw-Hill.
- Mills, L. B. (2009). A Meta-Analysis of the Relationship Between Emotional Intelligence and Effective Leadership. *Journal of Curriculum and Instruction, 3*(2), 22. doi:10.3776/joci.%y.323622

- Morgeson, F. P., Mumford, T. V., & Campion, M. A. (2005). Coming Full Circle: Using Research and Practice to Address 27 Questions About 360-Degree Feedback Programs. *Consulting Psychology Journal: Practice and Research*, 57(3), 196-209. doi:10.1037/1065-9293.57.3.196
- Moshavi, D., Brown, F. W., & Dodd, N. G. (2003). Leader self-awareness and its relationship to subordinate attitudes and performance. *Leadership and Organizational Development*, 24, 407-418.
- Mott, P. E. (1972). *The characteristics of effective organizations*. New York: Harper & Row.
- Mount, M. K. (1984). Psychometric proprieties of subordinate ratings of managerial performance. *Personnel Psychology*, 37(4), 687-702. doi:Article
- Mowday, R. T., Steers, R. M., & Porter, L. W. (1979). The measurement of organizational commitment. *Journal of Vocational Behavior*, 14(2), 224-247. doi:10.1016/0001-8791(79)90072-1
- Murphy, A. J. (1941). A Study of the Leadership Process. *American Sociological Review*, 6(5), 674. doi:10.2307/2085506
- Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, Organisational and Goal-based perspectives*. Thousand Oaks, CA: Sage Publications.
- Myers, I. B. (1962). *Manual : The Myers-Briggs Type indicator*. Princeton, NJ: Educational Testing Service.
- Myers, I. B., & McCaulley, M. H. (1985). *Manual: A guide to the development and use of the Myers-Briggs Type Indicator*. Palo Alto, CA: Consulting Psychologists Press.
- Myers, J. H., & Warner, W. G. (1968). Semantic Properties of Selected Evaluation Adjectives. *Journal of Marketing Research*, 5(4), 409-412.
- Narayanan, S., Venkatachalam, R., & Bharathiar, U. (1982). Leadership effectiveness and adaptability among small joisery units. *Managerial Psychology*, 3, 40-47.
- Newcombe, M. J., & Ashkanasy, N. J. (2002). The role of affect and affective congruence

in perceptions of leaders: an experimental study. *The Leadership Quarterly*, 13(5), 601-614. doi:10.1016/S1048-9843(02)00146-7

Nielsen, F., Fernandez, R. M., & Peng, S. S. (1982). *Hispanic students in American high schools: Background characteristics and achievement*. (National Center for Education Statistics). Washington, D.C.

Nordvik, H., & Brovold, H. (1998). Personality traits in leadership tasks. *Scandinavian Journal of Psychology*, 39(1), 61-64.

Norman, W. X. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, 66, 574-583.

Northouse, P. G. (2004). *Leadership: theory and practice*. (3rd Ed.). Thousand Oaks, CA: Sage Publications.

Northouse, P. G. (2007). *Leadership: theory and practice*. (4th edition.). Thousand Oaks, CA: Sage Publications.

Ostroff, C., Atwater, L. E., & Feinberg, B. J. (2004). Understanding self-other agreement: a look at rater and ratee characteristics, context and outcomes. *Personnel Psychology*, 57(2), 333-375. doi:Article

Patterson, C., Fuller, J. B., Kester, K., & Stringer, D. Y. (1995). *A meta-analytic examination of leadership style and selected compliance outcomes*. (Paper presented to the society for Industrial and Organizational Psychology). Orlando, FL.

Petty, M. M., & Miles, R. H. (1976). Leader sex-role stereotyping in a female-dominated work culture. *Personnel Psychology*, 29, 393-404.

Ployhart, R. E., Lim, B.-C., & Chan, K.-Y. (2001). Exploring relations between typical and maximum performance ratings and the five factor model of personality. *Personnel Psychology*, 54(4), 809-843. doi:Article

Podsakoff, P. M., MacKenzie, S. B., Moorman, R. H., & Fetter, R. (1990). Transformational leader behaviors and their effects on followers' trust in leader, satisfaction, and organizational citizenship behaviors. *The Leadership Quarterly*, 1(2), 107-142.

doi:10.1016/1048-9843(90)90009-7

- Ponnu, C., & Tennakoon, G. (2009). The Association Between Ethical Leadership and Employee Outcomes a the Malaysian Case. *Electronic Journal of Business Ethics and Organization Studies*, 14(1), 21-32.
- Prati, L. M., Douglas, C., Ferris, G. R., Ammeter, A. P., & Buckley, M. R. (2003). Emotional Intelligence, leadership effectiveness and team outcomes. *The international journal of organizational analysis*, 11, 21-40.
- Radman, H. (1964). An industrial use of peer ratings. *Journal of Applied Psychology*, 48(4), 211-214.
- Ragins, B. R. (1989). Power and Gender Congruency Effects in Evaluations of Male and Female Managers. *Journal of Management*, 15(1), 65. doi:Article
- Rasbash, J., Charlton, C., Browne, W. J., Healy, M., & Cameron, B. (2009). *MLwiN Version 2.1*. University of Bristol.: Centre for Multilevel Modelling.
- Rauch, C. F., & Behling, O. (1984). Functionalism : Basis for an alternate approach to the study of leadership. In J. G. Hunt, D. M. Hosking, C. A. Schriesheim, & R. Stewart (Éd.), *Leaders and managers: International perspectives on managerial behaviour and leadership* (p. 42-62). Elmsford: Pergamon Press.
- Reddin, W. J. (1986). The 3-D management style theory. *Training and Development Journal*, 21(4), 39-41.
- Reddin, W. J. (1987). *How to Make Management Style More Effective*. Maidenhead: McGraw Hill.
- Reilly, R. R., Smither, J. W., & Vasilopoulos, N. L. (1996). A longitudinal study of upward feedback. *Personnel Psychology*, 49(3), 599–612.
- Rice, R. W. (1978a). Psychometric Properties of the Esteem for Least Preferred Coworker (LPC Scale). *Academy of Management Review*, 3(1), 106-118. doi:10.5465/AMR.1978.4296576
- Rice, R. W. (1978b). Construct validity of the Least Preferred Co-Worker score.

Psychological Bulletin, 85(6), 1199-1237. doi:10.1037/0033-2909.85.6.1199

- Richards, D., & Engle, S. (1986). After the vision: Suggestions to corporate visionaires and vicion champions. In J. D. Adams (Éd.), *Transforming Leadership* (p. 199-215). Alexandria, VA: Miles River Press.
- Riordan, C. M. (2000). Relational demography within groups: Past developments, contradictions, and new directions, *Research in Personnel and Human Resources Management*(19), 131-173. doi:10.1016/S0742-7301(00)19005-X
- Roberson, L., & Block, C. J. (2001). Racioethnicity and job performance: A review and critique of theoretical perspectives on the causes of group differences. In B. M. Staw & R. I. Sutton (Éd.), *Research in organizational behavior* (Vol. 23, p. 247-325). Oxford, UK: JAI Press/Elsevier Science.
- Robie, C., Johnson, K. M., Nilsen, D., & Hazucha, J. F. (2001). The right stuff: understanding cultural differences in leadership performance. *Journal of Management Development*, 20(7), 639-650. doi:10.1108/EUM0000000005637
- Rowold, J., & Kersting, M. (2008). The Assessment of Charismatic Leadership. *European Journal of Psychological Assessment*, 24(2), 124-130. doi:10.1027/1015-5759.24.2.124
- Rubin, D. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley and Sons.
- Rubin, R. S., Munz, D. C., & Bommer, W. H. (2005). Leading from within: the effects of emotion recognition and personality on transformational leadership behavior. *Academy of Management Journal*, 48(5), 845-858. doi:10.5465/AMJ.2005.18803926
- Rushton, J. P., Brainerd, C. J., & Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin*, 94, 18-38.
- Sackett, P. R., DuBois, C. L., & Noe, A. W. (1991). Tokenism in performance evaluation: The effects of work group representation on male-female and White-Black differences in performance ratings. *Journal of Applied Psychology*, 76(2), 263-267.

doi:10.1037/0021-9010.76.2.263

- Sadler, P. J. (1970). Leadership Style, Confidence in Management, and Job Satisfaction. *The Journal of Applied Behavioral Science*, 6(1), 3-19. doi:10.1177/002188637000600101
- Sala, F. (2003). Executive Blind Spots: Discrepancies Between Self- and Other-Ratings. *Consulting Psychology Journal: Practice and Research*, 55(4), 222-229. doi:10.1037/1061-4087.55.4.222
- Sashkin, M. (2004). Transformational leadership approaches: A review and synthesis. In J. Antonakis, A. T. Cianciolo, & R. J. Sternberg (Éd.), *The nature of leadership* (p. 171– 196). Thousand Oaks, CA: Sage Punlication Inc.
- Schriesheim, C. A., Castro, S. L., & Cogliser, C. C. (1999). Leader-member exchange (LMX) research: A comprehensive review of theory, measurement, and data-analytic practices. *The Leadership Quarterly*, 10(1), 63-113. doi:10.1016/S1048-9843(99)80009-5
- Schriesheim, C. A., & Murphy, C. J. (1976). Relationships between leader behavior and subordinate satisfaction and performance: A test of some situational moderators. *Journal of Applied Psychology*, 61(5), 634-641. doi:10.1037/0021-9010.61.5.634
- Schwarz, J. C., Barton-Henry, M. L., & Pruzinsky, T. (1985). Assessing Child--rearing Behaviors: A Comparison of Ratings Made by Mother, Father, Child, and Sibling on the CRPBI. *Child Development*, 56(2), 462. doi:Article
- Seeman, M. (1963). Alienation and Social Learning in a Reformatory. *American Journal of Sociology*, 69(3), 270-284.
- Seeman, M., & Evans, J. W. (1962). Alienation and learning in a hospital setting. *American Sociological Review*, 27, 772-783.
- Sevy, B. A., Olson, R. D., McGuire, D. P., Frazier, M. E., & Paaganen, G. (1985). *Management skills profile technical manual*. Minneapolis, MN: Personnel Decisions.
- Shackleton, V. J. (1995). *Business Leadership*. London: Routledge.

- Shamir, B. (1991). Meaning, self and motivation in organizations. *Organization Studies*, 12, 405-424.
- Shamir, B., House, R. J., & Arthur, M. B. (1993). The motivational effects of charismatic leadership: a self-concept based theory. *Organization Science*, 4(4), 577-594. doi:Article
- Shortell, S. M., Rousseau, D. M., Gillies, R. R., Devers, K. J., & Simons, T. L. (1991). Organizational Assessment in Intensive Care Units (ICUs): Construct Development, Reliability, and Validity of the ICU Nurse-Physician Questionnaire. *Medical Care August 1991*, 29(8), 709-726.
- Shrauger, J. S., & Schoeneman, T. J. (1979). Symbolic interactionist view of self-concept: Through the looking glass darkly. *Psychological Bulletin*, 86(3), 549-573.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420-428. doi:10.1037/0033-2909.86.2.420
- Skipton, L. H. (2003). Leadership development for the postindustrial, postmodern information age. *Consulting Psychology Journal: Practice and Research*, 55(1), 3-14. doi:10.1037/1061-4087.55.1.3
- Smither, J. W., London, M., Vasilopoulos, N. L., Reilly, R. R., Millsap, R. E., & Savemini, N. (1995). An examination of the effects of an upward feedback program over time. *Personnel Psychology*, 48(1), 1-34.
- Sparks, J. R., & Schenk, J. A. (2001). Explaining the effects of transformational leadership: an investigation of the effects of higher-order motives in multilevel marketing organizations. *Journal of Organizational Behavior*, 22(8), 849-869.
- Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- SPSS 17 for Windows. (2008).Rel. 17.0.0. Chicago: SPSS Inc.
- Srivastava, K. B., & Bharamanaikar, S. R. (2004). Emotional intelligence and effective leadership behaviour. *Psychological Studies*, 49(2-3), 107-113.

- Staw, B. M., Sutton, R. I., & Pelled, L. H. (1994). Employee Positive Emotion and Favorable Outcomes at the Workplace. *Organization Science*, 5(1), 51-71. doi:Article
- Stogdill, R. M. (1948). Personal factors associated with leadership: A survey of the literature. *Journal of psychology*, 25, 35-71.
- Stogdill, R. M. (1950). Leadership, membership and organization. *Psychological Bulletin*, 47(1), 1-14. doi:10.1037/h0053857
- Stogdill, R. M. (1962). *Leaders Behavior Description Questionnaire - Form XII: an experimental revision*. Ohio State University, Columbus, OH: Fisher College of Business.
- Stogdill, R. M. (1969). Validity of Leader Behavior descriptions. *Personnel Psychology*, 22, 153-158.
- Stogdill, R. M. (1974). *Handbook of leadership*. New York: Free Press.
- Tannenbaum, R., & Schmidt, W. H. (1958). How to choose a leadership pattern. *Harvard Business Review*, 36(2), 95-101.
- Tannenbaum, R., Weschler, I. R., & Massarik, I. (1961). *Leadership organization: A behavioral approach*. New York: McGraw-Hill.
- Tein, J. Y., Roosa, M. W., & Michaels, M. (1994). Agreement between Parent and Child Reports on Parental Behaviors. *Journal of Marriage and the Family*, 56(2), 341-355.
- Tejeda, M. J., Scandura, T. A., & Pillai, R. (2001). The MLQ revisited: psychometric properties and recommendations. *The Leadership Quarterly*, 12(1), 31-52. doi:10.1016/S1048-9843(01)00063-7
- Thomas, J. L., Dickson, M. W., & Bliese, P. D. (2001). Values predicting leader performance in the U.S. Army Reserve Officer Training Corps Assessment Center: evidence for a personality-mediated model. *The Leadership Quarterly*, 12(2), 181-196. doi:10.1016/S1048-9843(01)00071-6
- Tornow, W. W. (1993). Perceptions or reality: Is multi-perspective measurement a means

or an end? *Human Resource Management*, 32(2-3), 221-229.
doi:10.1002/hrm.3930320203

Tornow, W. W., & London, M. (1998). *Maximizing the value of 360 degree feedback, a process for successful individual and organizational development*. San Francisco, CA: Center for Creative Leadership.

Trevino, L. K., & Weaver, G. R. (2003). *Managing ethics in businesses organizations: Social scientific perspectives*. Stanford, CA: Stanford University Press.

Trice, H. M., & Beyer, J. M. (1986). Charisma and its routinization in two social movement organizations. *Research in Organizational Behavior*, 8, 113-164.

Tsui, A. S., & Gutek, B. A. (1984). A role set analysis of gender differences in performance, affective relationships, and career success of industrial middle managers. *Academy of Management Journal*, 27, 619-635.

Tsui, A. S., & O'Reilly, C. A. (1989). Beyond simple demographic effects: The importance of relational demography in superior-subordinate dyads. *Academy of Management Journal*, 32(2), 402-423. doi:10.2307/256368

Tsui, A. S., & Ohlott, P. (1988). Multiple assessment of managerial effectiveness: Interrater agreement and consensus in effectiveness models. *Personnel Psychology*, 41(4), 779-803.

Tupes, E. C., & Christal, R. E. (1961). *Recurrent personality factors based on trait ratings*. (Tech. Rep. No. ASD-TR-61-97). TX: U. S. Air Force: Lackland Air Force Base.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods*, 3(1), 4-70.

Vandenberg, R. J., & Self, R. M. (1993). Assessing newcomers' changing commitments to the organization during the first 6 months of work. *Journal of Applied Psychology*, 78(4), 557-568. doi:10.1037/0021-9010.78.4.557

Vandenberghe, C., Stordeur, S., & D'hoore, W. (2002). Transactional and Transformational Leadership in Nursing: Structural Validity and Substantive Relationships. *European*

Journal of Psychological Assessment, 18(1), 16-29. doi:10.1027//1015-5759.18.1.16

Vecchio, R. (1979). A Test of the Cognitive Complexity Interpretation of the Least Preferred Coworker Scale. *Educational and Psychological Measurement*, 39(3), 523-526. doi:10.1177/001316447903900301

Vecchio, R., & Anderson, R. J. (2009). Agreement in Self-Other Ratings of Leader Effectiveness: The role of demographics and personality. *International Journal of Selection and Assessment*, 17(2), 165-179. doi:10.1111/j.1468-2389.2009.00460.x

Visser, B. A., Ashton, M. C., & Vernon, P. A. (2008). What Makes You Think You're so Smart? *Journal of Individual Differences*, 29(1), 35-44. doi:10.1027/1614-0001.29.1.35

Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2002). The moderating influence of job performance dimensions on convergence of supervisory and peer ratings of job performance: Unconfounding construct-level convergence and rating difficulty. *Journal of Applied Psychology*, 87(2), 345-354. doi:10.1037/0021-9010.87.2.345

Vrignaud, P. (2006). La mesure de la litt ratie dans PISA: La m thodologie est la r ponse, mais quelle  tait la question? *Revue fran aise de p dagogie*, 157, 27-41.

Vrignaud, P., & Loarer, E. (2008). Tests et recrutement. In S. Ionescu & A. Blanchet ( d.), *Nouveau cours de psychologie* (Vol. Psychologie sociale et ressources humaines, coordonn  par M. Bromberg et A. Trognon, p. 331-364). Paris: PUF.

Vroom, V. H. (2003). Educating Managers for decision making and leadership. *Management Decision*, 41(10), 968-979.

Vroom, V. H., & Jago, A. G. (1988). *The new leadership: Managing participation in organizations*. Englewood Cliffs, NJ: Prentice-Hall.

Vroom, V. H., & Yetton, P. W. (1973). *Leadership and decision-making*. Pittsburgh: University of Pittsburgh Press.

Waldman, D. A., Bass, B. M., & Yammarino, F. J. (1990). Adding to contingent-reward behavior: The augmenting effects of charismatic leadership. *Group and*

- Organization Studies*, 15(4), 381-394.
- Waldman, D., Atwater, L. E., & Antonini, D. (1998). Has 360 degree feedback gone amok? *Academy of Management Executive*, 12, 86-94.
- Walker, A. G., & Smither, J. W. (1999). A five-year study of upward feedback: What managers do with their results matters. *Personnel Psychology*, 52(2), 393–423.
- Watkins, D., & Cheung, S. (1995). Culture, gender, and response bias: An analysis of responses to the Self-Description questionnaire. *Journal of Cross-Cultural Psychology*, (26), 490-504.
- Weber, M. (1948). *From Max Weber: Essays in Sociology*. (H. Gerth & C. Wright Mills, Éd.). London: Routledge & Kegan Paul.
- Weber, M. (1968). *Economy and Society*. (G. Roth & C. Wittich, Éd.) (1925^e éd., Vol. 3). New York: Bedminster.
- Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of personality and social psychology*, 39(5), 806.
- Westburgh, E. M. (1931). A point of view—studies in leadership. *The Journal of Abnormal and Social Psychology*, 25, 418-423.
- Whetten, D. A., & Cameron, K. S. (2007). *Developing management skills*. Upper Saddle River, NJ: Prentice Hall.
- Wohlers, A. J., & London, M. (1989). Ratings of managerial characteristics : Evaluation difficulty, co-worker agreement, and self-awareness. *Personnel Psychology*, 42(2), 235-261. doi:Article
- Wong, C.-S., & Law, K. S. (2002). The effects of leader and follower emotional intelligence on performance and attitude. *The Leadership Quarterly*, 13(3), 243-274. doi:10.1016/S1048-9843(02)00099-1
- Xie, J. L., Roy, J.-P., & Chen, Z. (2006). Cultural and Individual Differences in Self-Rating Behavior: An Extension and Refinement of the Cultural Relativity Hypothesis. *Journal of Organizational Behavior*, 27(3), 341-364.

- Yammarino, F. J., & Atwater, L. E. (1997). Do Managers See Themselves as Others See Them? Implications of Self-Other Rating Agreement for Human Resources Management. *Organizational Dynamics*, 25(4), 35-44. doi:Article
- Yammarino, F. J., & Atwater, L. E. (2001). Understanding agreement in multisource feedback. In D. Bracken, C. Timmreck, & A. Church (Éd.), *The handbook of multisource feedback*. (p. 204-220). San Francisco, CA: Jossey-Bass.
- Yammarino, F. J., Dionne, S. D., Uk Chun, J., & Dansereau, F. (2005). Leadership and levels of analysis: A state-of-the-science review. *The Leadership Quarterly*, 16(6), 879–919.
- Yik, M. S. M., Bond, M. H., & Paulhus, D. L. (1998). Do Chinese Self-Enhance or Self-Efface? It's a Matter of Domain. *Personality and Social Psychology Bulletin*, 24(4), 399-406. doi:10.1177/0146167298244006
- Yukl, G. (1994). *Leadership in organizations*. (3rd edition.). Englewood Cliffs, NJ: Prentice Hall.
- Yukl, G. (2002). *Leadership in organizations*. (5th edition.). Upper Saddle River, NJ: Prentice Hall.
- Yukl, G., Gordon, B., & Taber, T. (2002). A Hierarchical Taxonomy of Leadership Behavior: Integrating a Half Century of Behavior Research. *Journal of leadership and organizational studies*, 9, 15-32.
- Yukl, G., & Tracey, J. B. (1992). Consequences of influence tactics used with subordinates, peers, and the boss. *Journal of Applied Psychology*, 77(4), 525.
- Yukl, G., Wall, S., & Lepsinger, R. (1990). Preliminary report on validation of the managerial practices survey. In K. E. Clark & M. B. Clark (Éd.), *Measures of leadership* (p. 223-238). West Orange, NJ: Leadership Library of America.
- Zagorsek, H., Stough, S. J., & Jaklic, M. (2006). Analysis of the Reliability of the Leadership Practices Inventory in the Item Response Theory Framework. *International Journal of Selection and Assessment*, 14(2), 180-191. doi:10.1111/j.1468-2389.2006.00343.x

Annexe

Mariya Jilinskaya

Thèse soutenue en octobre 2012, Université Paris Ouest Nanterre La Défense

Approche psychométrique et différentielle de la mesure du leadership par la méthode à 360 degrés.

Artefact et réalité dans l'hétéro-évaluation.

Résumé

Cette thèse porte sur l'évaluation du leadership par une approche multi-évaluateurs (dite à 360 degrés). Tout d'abord les différents modèles du leadership, allant d'une conception unitaire, à une conception interactionniste puis à une définition en termes d'effet, sont détaillés. Puis en étudiant la question de la mesure, on met en évidence qu'avec la popularité croissante de leurs modèles, certains outils d'évaluation sont devenus des questionnaires psychométriques à part entière. Pourtant, du fait des limites de l'auto-évaluation, une nouvelle approche du leadership à vu le jour: l'évaluation à 360°. Elle évalue les qualités d'un manager en interrogeant les personnes travaillant avec lui (subordonnés, collègues, supérieurs...) et en comparant leurs évaluations avec la propre évaluation du manager. Un des points central de notre recherche a été d'étudier les apports et les limites de cette méthode. Tout d'abord on a vérifié dans quelle mesure les caractéristiques souvent utilisées pour expliquer la variabilité entre les catégories d'observateurs permettaient réellement de comprendre les écarts observés. Ces analyses ont montré que malgré des résultats significatifs, ces variables n'expliquent que très partiellement la variance existante. De par ces conclusions l'accent a été mis, non plus sur les différences inter-groupes, mais sur l'accord et le désaccord au sein des groupes d'observateurs. Enfin, la dernière partie revient aux bases méthodologiques et théoriques de la mesure en cherchant à proposer un modèle psychométriques qui conviendrait aux résultats de questionnaires à 360°, permettant de donner un cadre conceptuel au recours à des évaluateurs multiples.

Psychometric and differential approach to leadership assessment with 360 degree.

Artifact and reality of interrater agreement

Abstract

This thesis is centered on leadership assessment through multi-rater evaluation, commonly known as 360 degrees assessment. First, leadership models were presented, and then, we discussed the measurement aspects of leadership, wherein we observed that some tools became full fledged psychometric assessments owing to the growing popularity of their underlying theory. Nevertheless, the concerns over the inherent limitations of self-report measures continued to be a major challenge in leadership assessment. This led to a new assessment approach called 360 degrees in which the characteristics of leaders are assessed by people working with them (subordinates, colleagues, superiors..) and compared with the leaders' self-appraisals. The focal point of this thesis was to study the advantages as well as the limitations of this approach. The study started with examining how well the variables which are supposed to explain the inter-rater variability were actually helpful in understanding the observed variance among observers. Those analysis yielded significant results despite the fact that those variables could explain only a very limited amount of variance. Following these observations, the study switched its focus from inter-group differences to intra-group / inter-rater agreement and disagreement. Finally, the last part of this thesis gets back to methodological and theoretical basics of measurement theory and proposes a psychometric model that would suit the 360 degrees assessments followed by a conceptual framework for the studies using multi-rater techniques.

Mots clefs : Leadership, 360 degrés, analyses multi-niveaux, accord inter-observateur, psychométrie, variabilité inter et intra-individuelle

Key words : Leadership, 360 degrees, multi-level analysis, inter-rater agreement, psychometrics, inter and intra-individual variability.